# THIRD-GENERATION TB-LMTO

O.K. ANDERSEN, C. ARCANGELI, R.W. TANK,
T. SAHA-DASGUPTA, G. KRIER, O. JEPSEN, and I. DASGUPTA.
Max-Planck Institut für Festkörperforschung, Stuttgart, Germany

ABSTRACT

We describe the screened Korringa-Kohn-Rostoker (KKR) method and the third-generation linear muffin-tin orbital (LMTO) method for solving the single-particle Schrödinger equation for a MT potential. In the screened KKR method, the eigenvectors $c_{RL,i}$ are given as the non-zero solutions, and the energies $\varepsilon_i$ as those for which such solutions can be found, of the linear homogeneous equations: $\sum_{RL} K^a_{R'L',RL}(\varepsilon_i)\, c_{RL,i} = 0$, where $K^a(\varepsilon)$ is the screened KKR matrix. The screening is specified by the boundary condition that, when a screened spherical wave $\psi^a_{RL}(\varepsilon, \mathbf{r}_R)$ is expanded in spherical harmonics $Y_{R'L'}(\hat{\mathbf{r}}_{R'})$ about its neighboring sites $\mathbf{R}'$, then each component either vanishes at a radius, $r_{R'}$=$a_{R'L'}$, or is a regular solution at that site. When the corresponding "hard" spheres are chosen to be nearly touching, then the KKR matrix is usually short ranged and its energy dependence smooth over a range of order 1 Ry around the centre of the valence band. The KKR matrix, $K(\varepsilon_\nu)$, at a fixed, arbitrary energy turns out to be the negative of the Hamiltonian, and its first energy derivative, $\dot{K}(\varepsilon_\nu)$, to be the overlap matrix in a basis of kinked partial waves, $\Phi_{RL}(\varepsilon_\nu, \mathbf{r}_R)$, each of which is a partial wave inside the MT-sphere, tailed with a screened spherical wave in the interstitial, or taking the other point of view, a screened spherical wave in the interstitial, augmented by a partial wave inside the sphere. When of short range, $K(\varepsilon)$ has the two-centre tight-binding (TB) form and can be generated in real space, simply by inversion of a positive definite matrix for a cluster. The LMTOs, $\chi_{RL}(\varepsilon_\nu)$, are smooth orbitals constructed from $\Phi_{RL}(\varepsilon_\nu, \mathbf{r}_R)$ and $\dot{\Phi}_{RL}(\varepsilon_\nu, \mathbf{r}_R)$, and the Hamiltonian and overlap matrices in the basis of LMTOs are expressed solely in terms of $K(\varepsilon_\nu)$ and its first *three* energy derivatives. The errors of the single-particle energies $\varepsilon_i$ obtained from the Hamiltonian and overlap matrices in the $\Phi(\varepsilon_\nu)$- and $\chi(\varepsilon_\nu)$ bases are respectively of second and fourth order in $\varepsilon_i - \varepsilon_\nu$. Third-generation LMTO sets give wave functions which are correct to order $\varepsilon_i - \varepsilon_\nu$, not only inside the MT spheres, but also in the interstitial region. As a consequence, the simple and popular formalism which previously resulted from the atomic-spheres approximation (ASA) now holds in general, that is, it includes downfolding and the combined correction. Downfolding to few-orbital, possibly short-ranged, low-energy, and possibly orthonormal Hamiltonians now works exceedingly well, as is demonstrated for a high-temperature superconductor. First-principles $sp^3$ and $sp^3d^5$ TB Hamiltonians for the valence and lowest conduction bands of silicon are derived. Finally, we prove that the new method treats overlap of the potential wells correctly to leading order and we demonstrate how this can be exploited to get rid of the empty spheres in the diamond structure.

## INTRODUCTION

There is a need for an *intelligible* and *accurate* first-principles electronic-structure method. Our efforts have been directed towards developing a single-particle basis which, for the application at hand, can be adjusted to a useful compromise between being *short ranged, minimal,* and *accurate.*

### Recent developments of Multiple Scattering Theory

Since atoms are nearly round, it seems most natural to start out using *spherical waves,* $j_l(\kappa r)\,Y_L(\hat{\mathbf{r}})$ and $n_l(\kappa r)\,Y_L(\hat{\mathbf{r}})$, where $L{=}lm$, as done when solving Schrödinger's equation with the classical multiple-scattering method due to Korringa, Kohn, and Rostoker (KKR).[1] In this method the scattering by the atom at site $\mathbf{R}$ is specified by the *phase shifts,* $\eta_{Rl}(\kappa)$, of its potential well, and the structure of the solid is specified by a Hermitian matrix with elements $B_{RL,R'L'}(\kappa)$, the *structure constants.* In terms of these, the wave-function coefficients, $c_{RL,i}$, are the solutions of the homogeneous, linear equations, one for each $R'L'$ :

$$\sum_{RL}\left[B_{R'L',RL}(\kappa)+\kappa\cot\eta_{Rl}(\kappa)\,\delta_{R'L',RL}\right]c_{RL,i}=0, \tag{1}$$

and the energies, $\varepsilon_i$, are the values of $\kappa^2$ $(\equiv\varepsilon)$ for which solutions can be found, *i.e.* the determinant of $B(\kappa)+\kappa\cot\eta(\kappa)$ vanishes. There are merely 4–16 equations per atom because all phase shifts with $l>1$–3 vanish. The KKR equations provide the *exact* solutions of Schrödinger's equation, but only for a muffin-tin (MT) potential, $V(\mathbf{r})\equiv\sum_R v_R(|\mathbf{r}-\mathbf{R}|)$, which is a superposition of spherically symmetric, non-overlapping potential wells, $v_R(r)$, of ranges $s_R$.

There are three problems with the KKR method: First of all, a non-overlapping MT potential is a poor representation of the self-consistent potential in any, except the most close packed solid. Secondly, the structure constants have long range, and thirdly, strong energy dependence. Specifically, the energy dependence of the KKR matrix, $B(\kappa)+\kappa\cot\eta(\kappa)$, is not linear like that of the secular matrix, $H-\varepsilon O$, for an energy eigenvalue problem in an energy-independent, possibly non-orthonormal representation. The main reason for the second and third drawbacks is that the spherical Bessel $(j_l)$ and Neumann $(n_l)$ functions have long range and depend on energy. This leads to interferences, which cause long range and strong energy dependence of the structure matrix. For a crystal with lattice translations $\mathbf{T}$, the Bloch-summed structure matrix, $B_{R'L',RL}(\kappa,\mathbf{k})\equiv\sum_T\exp(i\mathbf{k}\cdot\mathbf{T})\,B_{(R'+T)L',RL}(\kappa)$, must be evaluated by the Ewald procedure, has poles at the free-electron parabola, $\kappa^2=\sum_G|\mathbf{k}+\mathbf{G}|^2$, and a branch cut at the bottom of the continuum, $\kappa{=}0$.

It was recently shown,[2] and we shall present a slightly different proof below, that even when the potential wells *overlap,* the KKR equations do hold to first order in the potential overlap. This, as we shall demonstrate, allows the use of MT spheres with up to at least 50 per cent radial overlap [$s_R+s_{R'}\lesssim 1.5\,|\mathbf{R}-\mathbf{R}'|$ for $R$ and $R'$ denoting nearest neighbors], and hence treat the potential between the atoms in a more realistic way. With such large overlaps, the zero of the potential moves from the potential threshold between the atoms

towards the vacuum level, and this means that the energies for the occupied states are usually negative.

It was furthermore shown,[2] that transformation to *linear combinations* of spherical Bessel and Neumann functions, so-called *screened* spherical waves, characterized by a set of background phase shifts $\alpha_{Rl}(\kappa)$, can remove the long range *and* the strong energy dependence from the structure matrix, provided that the energy is not too high. This screening transformation may be expressed as:

$$|n^\alpha\rangle = |n\rangle \left(1 - \frac{\tan\alpha}{\kappa} B^\alpha\right), \text{ where } \tan\eta^\alpha = \tan\eta - \tan\alpha, \text{ and } [B^\alpha]^{-1} = B^{-1} + \frac{\tan\alpha}{\kappa}. \quad (2)$$

These are, respectively, vector-, scalar-, and matrix equations. The superscript labels the representation. In the first equation, we have used a notation in which $|n\rangle$ is a row vector of functions with components $n_l(\kappa r_R) Y_{lm}(\hat{\mathbf{r}}_R)$ and where $\mathbf{r}_R \equiv \mathbf{r} - \mathbf{R}$. The last equation involves inversion of the matrix $B + \kappa\cot\alpha$. For a set of background phase shifts, which are known to give short range, this inversion may be performed in real space and the screened KKR method is basically a first principles tight-binding (TB) method. At the time,[2] however, the relation between range and background phase shifts was poorly understood. This problem was solved later[3] by expressing the background phase shifts in terms of their hard-sphere radii, $a_{RL}$, defined by

$$\tan\alpha_{RL}(\kappa) \equiv j_l(\kappa a_{RL}) / n_l(\kappa a_{RL}), \quad (3)$$

or equivalently, by letting the background phase shifts be those of repulsive potential wells.[4] Looked upon in this way, the role of the confinement is to push the bottom of the continuum up in energy with respect to the floor of the MT potential, and thereby, to leave below a range of energies in which the confined wave-equation solutions are localized and which, in order to be useful, should include the range of the occupied bands for the real (attractive) potential. With the definition (2), the screened spherical waves, $|n^\alpha\rangle$, are still quite energy dependent, but only due to their normalization. A more suitable normalization followed naturally from the hard-sphere point of view.[3]

Finally, it was pointed out[2] that the screening transformation may also be used to remove unwanted channels from the KKR equations by choosing, for those channels, the background phase shifts equal to the real phase shifts, $\alpha(\kappa) = \eta(\kappa)$. This is a transformation to a *minimal basis.*

With these three recent developments, we have the basic ingredients for the Schrödinger part of an *intelligible and accurate* method, the third-generation LMTO method.

Earlier developments; the Atomic Spheres Approximation

The attempts to develop from the KKR method an *intelligible* first-principles method were initiated 25 years ago[5] and overlapping spheres, the two-centre interpretation, and screening transformations[6] have been used routinely for a long time. The new development,[3] which started six years ago,[2] and which will be further elaborated on in the present paper,

aims at making the method also *accurate* without loss of simplicity and elegance. Whereas in the earlier developments reduced range and energy dependence were achieved through a physically motivated approximation, namely the atomic-spheres approximation (ASA), this is now achieved *exactly,* and that in turn, allows a *controlled* approximation for the potential overlap.

The ASA[5,7] consists of letting the MT spheres overlap to the extent that they become space filling, whereby the interstitial region is effectively eliminated so that one may neglect the energy dependence of the wave functions in this region and, hence, the energy dependence of the structure matrix. The remaining energy dependence now occurs only along the diagonal of the KKR-ASA matrix where it enters through the radial logarithmic derivatives evaluated at the atomic sphere,

$$D\left\{\phi_{Rl}\left(\varepsilon, s_R\right)\right\} \equiv s_R \phi'_{Rl}\left(\varepsilon, s_R\right) / \phi_{Rl}\left(\varepsilon, s_R\right). \tag{4}$$

If, as is usually done, the kinetic energy in the interstitial is taken to be zero ($\kappa^2 = 0$), the suitably renormalized KKR-ASA equations become:

$$\sum_{RL}\left[\mathsf{S}^0_{R'L',RL} - \mathsf{P}^0_{Rl}\left(\varepsilon\right)\delta_{R'L',RL}\right] c_{RL,i} = 0, \quad \text{where}$$

$$\mathsf{P}^0_{Rl}\left(\varepsilon\right) \equiv 2\left(2l+1\right)\left(\frac{w}{s_R}\right)^{2l+1} \frac{D\left\{\phi_{Rl}\left(\varepsilon, s_R\right)\right\} + l + 1}{D\left\{\phi_{Rl}\left(\varepsilon, s_R\right)\right\} - l} \approx \left[\frac{\Delta_{Rl}}{\varepsilon - C_{Rl}} + \gamma_{Rl}\right]^{-1}$$

are the potential functions for well $v_R\left(r\right)$, and $C_{Rl}$, $\Delta_{Rl}$, and $\gamma_{Rl}$ are potential parameters. $\mathsf{S}^0$ is the structure matrix given by:

$$\mathsf{S}^0_{\text{on site}} = 0,\ \mathsf{S}^0_{ss\sigma} = -2\left(w/d\right),\ \mathsf{S}^0_{sp\sigma} = 2\sqrt{3}\left(w/d\right)^2,\ ,\ \mathsf{S}^0_{dd(\sigma,\pi,\delta)} = 10\left(w/d\right)^5\left(-6, 4, -1\right), \tag{5}$$

when we choose $\mathbf{R}' - \mathbf{R} \equiv \hat{\mathbf{z}}d$ and $w$ is an arbitrary length scale, usually chosen to be the average Wigner-Seitz radius. The structure matrix thus consists of effective hopping integrals. For monatomic crystals, this gave rise to the concept of canonical bands.[5] However, the $d^{-l-l'-1}$-decay of the hopping integral between orbitals of angular-momentum characters $l$ and $l'$ is too slow for a tight-binding scheme, except for $d$- and $f$-orbitals.

It was therefore a breakthrough when it became understood that similarity transformations could be performed on the KKR-ASA ($\kappa^2 = 0$) equations and could lead to short range.[6] The transformation from the bare to a screened representation, specified by screening constants $\mathsf{a}_{RL}$, is given by:

$$\mathsf{P}^{\mathsf{a}}\left(\varepsilon\right)^{-1} = \mathsf{P}^0\left(\varepsilon\right)^{-1} - \mathsf{a} \quad \text{and} \quad \left[\mathsf{S}^{\mathsf{a}}\right]^{-1} = \left[\mathsf{S}^0\right]^{-1} - \mathsf{a}, \tag{6}$$

which are respectively scalar- and matrix equations. $\mathsf{a}$ is a diagonal matrix with elements $\mathsf{a}_{Rl}$. The corresponding transformation for the resolvent, useful for Green-function and CPA calculations,[8] is:

$$\left[\mathsf{P}^{\mathsf{b}}\left(z\right) - \mathsf{S}^{\mathsf{b}}\right]^{-1} = \left(\mathsf{b} - \mathsf{a}\right)\frac{\mathsf{P}^{\mathsf{a}}\left(z\right)}{\mathsf{P}^{\mathsf{b}}\left(z\right)} + \frac{\mathsf{P}^{\mathsf{a}}\left(z\right)}{\mathsf{P}^{\mathsf{b}}\left(z\right)}\left[\mathsf{P}^{\mathsf{a}}\left(z\right) - \mathsf{S}^{\mathsf{a}}\right]^{-1}\frac{\mathsf{P}^{\mathsf{a}}\left(z\right)}{\mathsf{P}^{\mathsf{b}}\left(z\right)}. \tag{7}$$

This involves no matrix multiplications, but merely rescaling of matrix elements. As for TB theory, the screening transformation gave a formalism for the "kinetic" part of the often observed dependence of the hopping integrals and on-site elements on the environment.[9] The screening constants yielding short range were found empirically. The potential-dependent choice $\mathsf{a}_{Rl}$=$\gamma_{Rl}$, on the other hand, makes the energy dependence of the KKR-ASA matrix linear (to second order) so that $C + \sqrt{\Delta}\mathsf{S}^{\gamma}\sqrt{\Delta} \equiv h^{\gamma}$ becomes the Hamiltonian in an orthonormal, but not necessarily short-ranged basis. The transformation finally made it possible to remove channels from the KKR-ASA equations by choosing $\mathsf{a}_{Rl}\left(\varepsilon\right) = \mathsf{P}^{0}_{Rl}\left(\varepsilon\right)^{-1}$ for such channels.[10] This removal, or "downfolding", however, reintroduced energy dependence of the structure matrix.

When performing density-functional calculations one needs to solve not only Schrödinger's but also Poisson's equation, and with the ASA method this involves approximating not only the potential but also the charge density by a superposition of slightly overlapping, spherically symmetric contributions. This gives a very simple scheme which fails badly in describing total-energy changes caused by symmetry-lowering distortions,[11] however, *e.g.* the ASA can be used for calculation of pressure-volume relations,[12] but not for calculation of phonon frequencies. Moreover, since the potential spheres are supposed to be space filling in the ASA, open structures can only be treated if the interstices between the atoms are filled with "empty" spheres and this works well only for structures such as the diamond structure, where the interstices have high symmetry. Even in such a case, for the description to be intelligible all empty-sphere channels must be downfolded, and that introduces a rather strong, non-linear energy dependence of the structure matrix which cannot be treated in the LMTO-ASA approach to be discussed below.[10]

Linear Muffin-Tin Orbitals of the first and second generations

In practice one does not solve the KKR equations, but one uses Green functions in a short-ranged representation and at complex energies,[8] or one solves energy eigenvalue equations, $\sum_{RL}\left[H_{R'L',RL} - \varepsilon O_{R'L',RL}\right]c_{RL,i} = 0$, which are equivalent with the KKR equations in a certain energy range around some chosen energy, $\varepsilon_{\nu}$. In the linear muffin-tin orbital (LMTO) method,[13,5] such an eigenvalue problem is arrived at by using the Raleigh-Ritz variational principle for the Hamiltonian in a basis of LMTO's constructed from the radial Schrödinger-equation solutions, $\phi_{Rl}\left(\varepsilon, r\right)$, for the potential wells and their first energy derivatives, $\dot{\phi}_{Rl}\left(\varepsilon, r\right)$, *at* the chosen energy, $\varepsilon$=$\varepsilon_{\nu}$. In the interstitial region, the first and second generation LMTOs use the spherical waves at $\kappa^{2}_{\nu}$, but not their first energy derivatives, so that the energy dependence in the interstitial is suppressed. The second-generation LMTO formalism[6] is elegant, but only in the ASA and only if no channels have been downfolded. Under these conditions, the Hamiltonian and overlap matrices are expressed solely in terms of the structure matrix and the potential functions: The structure matrix enters the formalism in the form of a first-order, two-centre TB-Hamiltonian: $h \equiv \dot{\mathsf{P}}^{-1/2}\left(\mathsf{S} - \mathsf{P}\right)\dot{\mathsf{P}}^{-1/2}$, where as usual $\mathsf{P}\left(\varepsilon\right)$ is a diagonal matrix. Here and in the follow-

ing, the common superscript $\mathsf{a}$ is dropped and an omitted energy argument means that the energy is set to $\varepsilon_\nu$. In terms of this two-centre Hamiltonian the LMTO set may be expressed as $|\chi\rangle \equiv |\phi\rangle + \left|\dot{\phi}\right\rangle h$, a form which may be regarded as the matrix equivalent of the linear approximation $\phi_{Rl}(\varepsilon, r) \approx \phi_{Rl}(r) + \dot{\phi}_{Rl}(r)(\varepsilon - \varepsilon_\nu)$. In the basis of these LMTOs the Hamiltonian and overlap matrices are respectively:

$$\langle\chi\,|-\Delta + V - \varepsilon_\nu|\,\chi\rangle = h(1+oh) \quad \text{and} \quad \langle\chi \mid \chi\rangle = (1+ho)(1+oh) + hph, \tag{8}$$

where

$$o \equiv \left\langle \phi \mid \dot{\phi} \right\rangle = \frac{1}{2!}\ddot{\mathsf{P}}/\dot{\mathsf{P}} \quad \text{and} \quad p + o^2 \equiv \left\langle \dot{\phi}^2 \right\rangle = \frac{1}{3!}\dddot{\mathsf{P}}\,/\dot{\mathsf{P}} \tag{9}$$

are diagonal matrices and it has been assumed that $\phi_{Rl}(r_R)\,Y_{lm}(\hat{\mathbf{r}}_R)$ is normalized to unity in its sphere, *i.e.* that $\langle\phi^2\rangle = 1$. The overlap matrix is seen to be nearly factorized and one may therefore transform to the Löwdin-orthonormalized representation, $\left|\chi^\perp\right\rangle \equiv |\chi\rangle\,\langle\chi \mid \chi\rangle^{-1/2} \sim |\chi\rangle\,(1+oh)^{-1} = |\chi^\gamma\rangle$, in which one finds the following expansion for the Hamiltonian:

$$\left\langle\chi^\perp\,|-\Delta + V - \varepsilon_\nu|\,\chi^\perp\right\rangle = h - hoh + h\,[oho - (ph+hp)/2]\,h + \ldots\,. \tag{10}$$

When $\mathsf{a}$ gives short range this is a power series in a TB Hamiltonian, $h^{\mathsf{a}}$. Truncation of this series after the first term yields a spectrum which is accurate in an energy window of size $\sim (10o)^{-1} = \frac{1}{5}\dot{\mathsf{P}}/\ddot{\mathsf{P}}$ around $\varepsilon_v$, but distorted further away. Adding terms, increases the size of this window at the expense of including further hoppings. The form (10) has been useful in recursion calculations[14] for structurally disordered condensed matter.[15]

For a case like the diamond structure, where one only wants LMTOs centered on atoms, downfolding of the empty-sphere LMTOs is achieved by transformation of the structure matrix using: $\mathsf{a}_E = \mathsf{P}_E^0(\varepsilon_\nu)^{-1}$, with $E$ referring to the empty-sphere channels. The energy is here set to $\varepsilon_\nu$ because in the LMTO-ASA formalism the structure matrix must be energy independent. Now, an atom-centered LMTO has a tail which extends into the empty spheres, and here, it is substituted by the corresponding partial waves. The atom-centered LMTO is therefore: $|\chi_A\rangle \equiv |\phi_A\rangle + \left|\dot{\phi}_A\right\rangle h_{AA} + |\phi_E\rangle\,\dot{h}_{EA}$, with $\dot{h}_{EA} \equiv \left[-\partial\mathsf{P}_E^0(\varepsilon)^{-1}/\partial\varepsilon|_{\varepsilon_\nu}\right]^{-1/2}\mathsf{S}_{EA}(\varepsilon_\nu)\,\dot{\mathsf{P}}_A^{-1/2}$. This is the way in which the energy dependence of $\mathsf{a}_E(\varepsilon)$ enters, but only to linear order. The overlap matrix $\langle\chi_A \mid \chi_A\rangle$ will now contain the term $\dot{h}_{AE}\dot{h}_{EA}$ involving $A-E-A$ hoppings, in addition to the terms in (8). This is clumsy and ruins the near factorization of the overlap matrix. With downfolding, the power-series expression (10) for the LMTO Hamiltonian in the Löwdin-orthonormalized basis does therefore not apply.

Most LMTO calculations include non-ASA corrections to the Hamiltonian and overlap matrices, such as the combined correction for the neglected integrals over the interstitial region and the neglected partial waves of high $l$. This brings in the first energy derivative of the structure matrix, $\dot{\mathsf{S}}$, in a way which makes the formalism clumsy.[5,9] Our current, second-generation LMTO code[16] is useful and quite accurate for calculating energy bands because it includes downfolding in addition to the combined correction,[17] but the underlying formalism is so complicated that we never tried to publish it. On the other hand, the combined

correction is often important, and so is downfolding because it is the only accurate means of avoiding "ghost bands". The reason for the lost elegance beyond the ASA is that, whereas the LMTO basis is complete to first order in $\varepsilon - \varepsilon_\nu$ *inside* the spheres, it is only complete to zeroth order in the *interstitial*. A compact formalism is therefore obtained only when the interstitial region is neglected, and that is what the ASA does, simply by substituting the MT spheres by space-filling spheres and neglecting the overlap errors. The proof that the KKR equations hold to leading order for overlapping potentials[2] does not apply to the LMTO-ASA formalism.

There *are* LMTO methods sufficiently *accurate* to provide *ab initio* structural energies and forces within density-functional theory.[18] For the reason mentioned above, the LMTOs for such methods[19] are defined with respect to non-overlapping potentials, and since there is considerable probability that a valence or conduction electron is in the interstitial region, *outside* atom-centered, non-overlapping spheres, an accurate basis has to include extra degrees of freedom to describe this region, empty-sphere orbitals centered at interstitial sites and/or atom-centered LMTOs with tails of different kinetic energies (multiple kappa-sets). Moreover, these methods do not use small and short-ranged representations. Finally, since a non-overlapping MT potential is a poor approximation to the self-consistent potential, these methods *must* include the matrix elements of the full potential. Hence, the formalisms are set up to provide final, numerical results and by themselves provide little insight.

Third-generation LMTOs

In this paper we shall modify the LMTO set without increasing its size, in such a way that it becomes complete to first order in the interstitial region too. This is a rather natural thing to do, once the screened spherical waves have been defined in terms of hard-sphere radii. For the MT Hamiltonian, including downfolding, we shall regain the simple formulas from the ASA, provided that $|\phi\rangle$, $h$, $o$, and $p$ are suitably redefined. The Hamiltonian and overlap matrices are now given *solely* in terms of the screened and renormalized KKR matrix, which we shall name $K(\varepsilon_\nu)$, and its first three energy derivatives, $\dot{K}(\varepsilon_\nu)$, $\ddot{K}(\varepsilon_\nu)$, and $\dddot{K}(\varepsilon_\nu)$; the potential parameters and the structure matrix do not occur individually as in the formalisms of the previous generations. Third-generation LMTOs[3] thus *do* satisfy the definition that they form a basis constructed to reproduce the wave functions, $\Psi_i(\mathbf{r})$, for a MT potential to linear order in the deviation of the single-particle energy, $\varepsilon_i$, from a freely chosen level, $\varepsilon_\nu$. That is, the error of the wave functions is of second order in $\varepsilon_i - \varepsilon_\nu$ and the error of the single-particle energy is then of fourth order. When we use potential wells that overlap, the wave functions will be correct to linear order in the potential overlap and the energy error will be of second order. As we shall demonstrate, this will remedy all shortcomings mentioned above for the previous LMTO generations.

We shall only be concerned with solving Schrödinger's equation in the present paper and leave our LMTO-like expansion of the charge density, solution of Poisson's equation, and evaluation of the total energy and forces for future papers.[20]

We start with a concise yet self-contained derivation of the screened KKR method, which will lead to suitably renormalized versions of Eq.s (2). Then we derive an expression for the error caused by using this method for potential wells which overlap, and find that the error is of second order in the overlap. The weak energy dependence and short range of the screened and renormalized KKR matrix, $K(\varepsilon)$, is exploited by using it to generate few-orbital, low-energy, possibly orthonormal and short-ranged Hamiltonians for a generic high-temperature superconductor (HTSC). Thereafter we derive the new LMTO method and demonstrate by application to free electrons that its energy errors are really of fourth order in $\varepsilon_i - \varepsilon_\nu$. The power and flexibility of the new method is demonstrated by deriving for the HTSC and for diamond-structured silicon various LMTO sets. Using non-orthogonal $sp^3$ sets for Si, we can get an accurate first-principles description of the valence *and* conduction bands if a 12th-nearest-neighbor range is allowed in the Hamiltonian and overlap matrices. With $\varepsilon_\nu$ chosen in the middle of the valence band, a 6th-n.n. $sp^3$-set suffices for an accurate description of the valence band and a reasonable description of the conduction band. In order to halve the number of matrix elements, even for a non-orthogonal basis, we use a formalism analogous to (8) where the off-diagonal elements of $o$ and $p$ have been neglected so that $h$ is the only matrix. With this simplification of the Hamiltonian and overlap matrices, retaining the 6th-n.n. $sp^3$ basis and the low $\varepsilon_\nu$, the description of the valence band remains good and merely the conduction band deteriorates. Finally, it is possible to limit the range to 3rd nearest neighbors provided that $d$-orbitals are included in the basis. In the last section, we demonstrate that not only for the KKR method, but also for the new LMTO method, the overlap error is of second order and that this can be exploited to get completely rid of the empty-sphere wells in the diamond structure.

## SCREENED SPHERICAL WAVES

We start by defining sets of solutions of the wave equation, $[\Delta + \varepsilon]\,\psi(\varepsilon, \mathbf{r}) = 0$, so-called screened-spherical-wave sets, $\{\psi^a_{RL}(\varepsilon, \mathbf{r} - \mathbf{R})\}$, which will serve as interstitial (envelope) functions for the basis that we shall use for solving Schrödinger's equation. The members of a screened-spherical-wave (SSW) set are obtained by letting $R$ run over all atomic sites and $L$ over all angular-momenta for which the scattering is strong. The set is labelled by the superscript $a$. Instead of defining $\psi^a_{RL}(\varepsilon, \mathbf{r}_R)$ as a specific linear combination of spherical Neumann functions like in Eq. (2), we specify it in terms of an inhomogeneous boundary condition which is illustrated in Fig.s 1 and 2 and is given as follows:

Concentric with each MT sphere, $R'$, we imagine a series of possibly coinciding *"hard"* spheres with radii $a_{R'L'}$. Now, $\psi^a_{RL}(\varepsilon, \mathbf{r}_R)$ is that solution of the wave equation whose $Y_{R'L'}(\hat{\mathbf{r}}_{R'})$ projection on the $R'L'$ sphere equals $\delta_{RL,R'L'}$, that is, 1 on its own sphere and 0 on all other spheres. We do not associate SSWs and hard spheres with *weakly-* and *non-*scattering channels. For such a channel, the $Y_{R'L'}(\hat{\mathbf{r}}_{R'})$ projection of the SSWs is defined to be a regular solution of the corresponding radial Schrödinger equation, that is, it matches onto the irregular wave-equation solution $j_{l'}(\kappa r_{R'}) - \tan\eta_{R'l'}(\kappa)\; n_{l'}(\kappa r_{R'})$, times some con-

stant, $c^a_{R'L',RL}(\varepsilon)$. The weakly- and non-scattering channels are thus parts of the SSW and will not enter the screened KKR- and LMTO matrices explicitly. All high-$l'$ channels are non-scatterers [$\tan\eta_{R'l'}(\kappa)=0$] due to the dominance of the centrifugal barrier. Empty spheres are examples of a weak scatterers. *Strong* scatterers are then, by definition, those channels with which we associate SSWs and hard spheres. Note that all SSWs in the set have the same boundary condition, except for the $\delta_{RL,R'L'}$. The SSW set, $\{\psi^a_{RL}(\varepsilon,\mathbf{r}_R)\}\equiv|RL\rangle\langle\mathbf{r}|$, may thus be considered as an unperturbed Green function in a hybrid representation.

Fig. 2 shows an SSW for the hypothetical case of only strong scattering. Weak- and non-scattering channels would have shown up as little tails extending into the two hard spheres. Such tails may be seen in Fig. 4 where the dashed curve is an SSW for Si. In this figure we have set the radial functions of the strongly scattering channels to zero inside the hard spheres and, defined in this way, $\psi^a_{RL}(\varepsilon,\mathbf{r}_R)$ jumps by the amount $Y_L(\hat{\mathbf{r}}_R)$ at its own hard sphere, $r_R$=$a_{RL}$, and has kinks at all hard spheres. Had we instead chosen to continue also the strongly scattering channels of the SSW into the hard spheres, the SSW would have been smooth, but diverging at the sites of the strongly scattering atoms, each radial part going as $j_{l'}(\kappa r_{R'})-\tan\alpha_{R'L'}(\kappa)\,n_{l'}(\kappa r_{R'})$. In order to get more feeling for SSWs, let us consider some limiting cases:

If we specify $a_{RL}=a\to 0$ for all channels, we obtain the *bare* spherical waves. These are

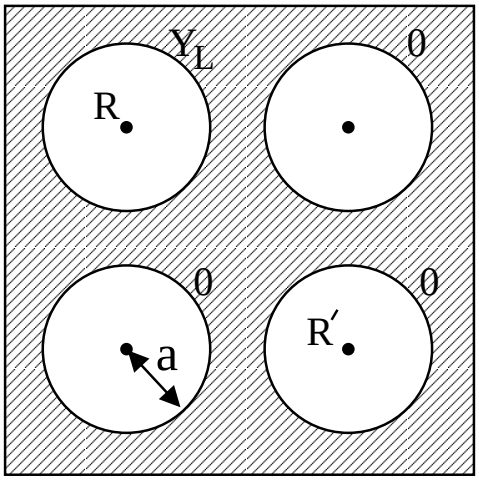


Figure 1: *(above).* Boundary condition for the screened spherical wave, $\psi^a_{RL}(\varepsilon,\mathbf{r}_R)$. Only strongly-scattering channels are indicated and all hard-sphere radii are equal.

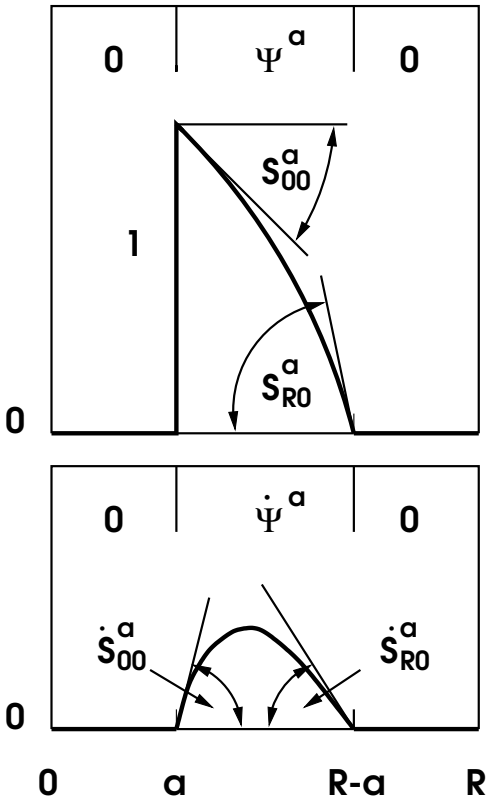


Figure 2: *(right).* Screened spherical wave centered at the origin, $\psi^a_0(\varepsilon,\mathbf{r})$, and its slopes, $S^a_{R,0}$ *(top).* The same for its first energy-derivative function, $\dot{\psi}^a_0(\varepsilon,\mathbf{r})$ *(bottom).*

known analytically but, except maybe for small molecules, we never use them. Nevertheless, with the normalization specified above they are:[3]

$$\begin{aligned}\psi^0_{0L}(\varepsilon,\mathbf{r}) &= -\frac{(\kappa a)^{l+1}}{(2l-1)!!}n_l(\kappa r)\,Y_L(\hat{\mathbf{r}}) = \left[\frac{a}{r}\right]^{l+1}\left[1+\frac{\varepsilon r^2}{2(2l-1)}-...\right]Y_L(\hat{\mathbf{r}}) \\ &\to \frac{(\kappa a)^{l+1}}{(2l-1)!!}\frac{\cos(\kappa r - l\pi/2)}{\kappa r}Y_L(\hat{\mathbf{r}}), \quad \text{for} \quad r\to\infty,\end{aligned}$$

when $\varepsilon \geq 0$, and where $(2l-1)!! \equiv (2l-1)(2l-3)..\cdot 1$ and $(-1)!! \equiv -1$. When $\varepsilon \leq 0$ :

$$\begin{aligned}\psi^0_{0L}(\varepsilon,\mathbf{r}) &= -\frac{(\kappa a)^{l+1}}{(2l-1)!!}\left[n_l(\kappa r)+ij_l(\kappa r)\right]Y_L(\hat{\mathbf{r}}) \\ &= \left[\frac{a}{r}\right]^{l+1}\left\{\left[1+\frac{\varepsilon r^2}{2(2l-1)}-\right]-\frac{(\varepsilon r^2)^l\, r\sqrt{-\varepsilon}}{(2l+1)\left[(2l-1)!!\right]^2}\left[1-\frac{\varepsilon r^2}{2(2l+3)}+\right]\right\}Y_L(\hat{\mathbf{r}}) \\ &\to \frac{a^{l+1}\left(\sqrt{-\varepsilon}\right)^l}{(2l-1)!!}\frac{\exp\left(-r\sqrt{-\varepsilon}\right)}{r}Y_L(\hat{\mathbf{r}}), \quad \text{for} \quad r\to\infty.\end{aligned}$$

These expressions hold only for $r \gg a$. Unlike screened spherical waves, the bare ones are eigenfunctions of angular momentum and are independent of the surroundings. For positive energies they have long range. Like all screened spherical waves, the normalization of the bare ones is such that they are dimensionless and, unlike the Bessel and Neumann functions, they depend little on energy near the hard spheres.

Another case is when we specify $a_{RL}$=$a_R$ for *all* $L$, and take $\varepsilon$=0. Then $\psi^a_{0L}(0,\mathbf{r})$ is proportional to the electrostatic potential from a $2^l$-pole at the origin, surrounded by grounded conducting spheres with radii $a_R$ centered at the other sites. Since the hard spheres at the neighbors break the spherical symmetry around the origin, the SSW has pure angular-momentum character merely at its own sphere, and this holds only as long as the own sphere coincides with all other spheres concentric with it. Changing the energy will not change the SSW much. If we now let $a_{R'L'}$ be zero for high $l'$'s, the SSWs will "wobble" into the hard spheres.

If the hard-sphere radii are generated from repulsive potential wells,[4] the SSWs are the "impurity states" for that repulsive MT potential.

Since the strongly-scattering components of the SSWs are forced to vanish at all surrounding hard spheres, the *range* of the SSWs depends on the choice of hard spheres and energy: Consider the spectrum $\varepsilon^a_i$ of the wave equation with the *homogeneous* boundary condition that the solutions vanish at *all* spheres. This spectrum has a continuum starting at $\varepsilon^a_c$, which in the absence of screening is at zero and which rises with increasing hard-sphere radii. Now, the SSWs are *localized* or *delocalized* depending on whether their energy is below or above the bottom of the continuum. Since we choose energy-independent boundary conditions for the SSWs, their *energy dependence* merely enters through the wave equation, that is through their curvature, and is therefore small when the wavelength exceeds the diameter of the largest interstitial in the hard-sphere solid.

If all hard spheres centered on the same site would coincide, then the hard spheres would have to be *smaller than touching* because, if two spheres had a point (or a circle) in common, then each one of the SSWs centered on the two spheres would be required to be both zero and non-zero at that point (or circle). When only a few low-$l$ channels scatter strongly, neighboring hard spheres may intersect. With decreasing hard-sphere interstitial, the SSW sets thus in general become more and more localized, until the hard spheres start to intersect. Since from there on, the SSWs are forced to change rapidly near the common circles, their behavior becomes chaotic as the circles grow.

We shall generate the screened spherical waves from the bare ones, because those are the only ones we know analytically. Hence, we first consider the question of how to expand an arbitrary wave-equation solution, $\Psi(\varepsilon, \mathbf{r})$, which is regular in all space, except possibly at the atomic sites, in an SSW set, $\{\psi^a_{RL}(\varepsilon, \mathbf{r}_R)\}$, with the same energy. If the number of atoms is finite, this energy is supposed to be negative. Moreover, since $\Psi(\varepsilon, \mathbf{r})$ is a solution of the wave equation, the SSW set is supposed to have no weakly-scattering channels and, *a priory,* we treat all channels as strong scatterers. Finally, we shall not truncate the SSWs inside the hard spheres, but let them continue to the centers. We now expand $\Psi(\varepsilon, \mathbf{r})$ in spherical harmonics on the hard spheres of the SSW set, thus obtaining the coefficients, $\Psi_{Rlm}(\varepsilon, a_{Rlm})$. Unless all of these vanish, the linear combination converges to $\Psi(\varepsilon, \mathbf{r})$ :

$$\lim_{\lambda_R \to \infty} \sum_R \sum_{l=0}^{\lambda_R} \sum_{m=-l}^{l} \psi^a_{Rlm}(\varepsilon, \mathbf{r}_R)\ \Psi_{Rlm}(\varepsilon, a_{Rlm}) = \Psi(\varepsilon, \mathbf{r})$$

because, by construction, the linear combination is a solution of the wave equation with the proper energy, and this solution matches $\Psi(\varepsilon, \mathbf{r})$ channel by channel. In order to convince oneself that the latter is sufficient, one may start repeating the argument using an SSW set with $L$-independent hard spheres. In that case, $\Psi(\varepsilon, \mathbf{r})$ coincides with the linear combination on a closed boundary, because in the case where the system is infinite such a boundary is formed by the entity of all hard spheres, and in the case where the system is finite, the boundary is formed by the hard spheres plus the infinity, where both $\Psi(\varepsilon, \mathbf{r})$ and the linear combination vanish since the energy is negative.

If all the coefficients $\Psi_{Rlm}(\varepsilon, a_{Rlm})$ vanish then $\Psi(\varepsilon, \mathbf{r})$ is an eigenfunction of the hard-sphere solid. In this case a complete set must include, in addition to the SSWs, the degenerate eigenfunctions, or we may choose a different SSW set for the expansion of $\Psi(\varepsilon, \mathbf{r})$.

Changing the hard-sphere radii, but not the sites and the energy, produces another set of SSW's which is also complete in the above-mentioned sense. All such sets are therefore linearly dependent. A set of hard-sphere radii is said to specify a *representation* and the transformation from the $a$ to the $b$ representation is obtained by substituting $\psi^b_{R'L'}(\varepsilon, \mathbf{r}_{R'})$ for $\Psi(\varepsilon, \mathbf{r})$ in the above. Hence, the transformation is

$$\psi^b_{R'L'}(\varepsilon, \mathbf{r}_{R'}) = \sum_{RL} \psi^a_{RL}(\varepsilon, \mathbf{r}_R)\ \psi^b_{RL,R'L'}(\varepsilon, a_{RL}), \tag{11}$$

where $\psi^b_{RL,R'L'}(\varepsilon, a_{RL})$ are the $RL$ components at $a_{RL}$-spheres of the functions $\psi^b_{R'L'}(\varepsilon, \mathbf{r}_{R'})$.

Expanding now the left and right-hand sides in spherical harmonics on the $b_{R''L''}$-spheres we obtain:

$$\delta_{R''R'}\delta_{L''L'} = \sum_{RL} \psi^a_{R''L'',RL}\left(\varepsilon, b_{R''L''}\right)\, \psi^b_{RL,R'L'}\left(\varepsilon, a_{RL}\right). \tag{12}$$

The two matrices $\psi^a\left(\varepsilon, b\right)$ and $\psi^b\left(\varepsilon, a\right)$ are thus each others inverses.

SLOPE AND STRUCTURE MATRICES

We have specified the SSWs by their nodes and shall need their radial derivatives at the hard spheres, that is, the dimensionless *slope matrix*. Its element $S^a_{R'L',RL}\left(\varepsilon\right)$ is defined as $a_{R'L'}$ times the $L'$-component of the radial derivative at the $a_{R'L'}$-sphere, with the positive direction taken outwards from $\mathbf{R}'$, of $\psi^a_{RL}\left(\varepsilon, \mathbf{r}_R\right)$. This is illustrated in Fig. 2.

In fact knowledge of the hard spheres and the slope matrix makes generation of the SSW set a simple matter: The spherical-harmonics expansion around any site, $\mathbf{R}'$, of any member, $\psi^a_{RL}\left(\varepsilon, \mathbf{r}_R\right)$, of the set is given by radial functions and the function for the $L'$ channel is:

$$\psi^a_{R'L',RL}\left(\varepsilon, r_{R'}\right) = \; f_{l'}\left(\varepsilon, a_{R'L'}, r_{R'}\right)\, \delta_{R'L',RL} + g_{l'}\left(\varepsilon, a_{R'L'}, r_{R'}\right)\, S^a_{R'L',RL}\left(\varepsilon\right). \tag{13}$$

The local expansion converges for $r_{R'}$ smaller than the distance to the nearest site. In (13), $f_l$ and $g_l$ are solutions of the radial wave equation, $\left[d^2/dr^2 - l\left(l+1\right)/r^2 + \varepsilon\right] r f_l = 0$, with the following boundary conditions for $r{=}a:$ $f$ has value one and slope zero, and $g$ has value zero and slope $1/a$.

The SSW-set, $\left|\psi^a\left(\varepsilon\right)\right\rangle$, may also be expressed *globally* as a linear combination of some known set, $\left|\psi^b\left(\varepsilon\right)\right\rangle$, as we saw in Eq. (11). The transformation matrix, $\psi^a\left(\varepsilon, b\right)$, is then given by Eq. (13) with $r_{R'}$ substituted by $b_{R'L'}$. With the use of Eq. (13), the completeness relation (12) thus expresses the transformation from $S^b\left(\varepsilon\right)$ to $S^a\left(\varepsilon\right)$.

In order to *generate* the slope matrix, we transform to the bare set, which is known analytically: Using Eq. (13), $S^a\left(\varepsilon\right)$ is expressed in terms of $\psi^a\left(\varepsilon, 0\right)$, which is computed as the inverse of $\psi^0\left(\varepsilon, a\right)$. The latter follows from the local, spherical-harmonics expansion about $\mathbf{R}'$ of the Neuman function centered at $\mathbf{R}$ $\left(\neq \mathbf{R}'\right)$: $\kappa n_l\left(\kappa \mathbf{r}_R\right) Y_L\left(\hat{\mathbf{r}}_R\right) = \sum_{L'} j_{l'}\left(\kappa \mathbf{r}_{R'}\right) Y_{L'}\left(\hat{\mathbf{r}}_{R'}\right) B_{R'L',RL}\left(\kappa\right)$, where

$$B_{R'L',RL}\left(\kappa\right) \equiv \sum_{l''} 4\pi\, i^{-l+l'-l''} C_{LL'l''}\, \kappa n_{l''}\left(\kappa\left|\mathbf{R}-\mathbf{R}'\right|\right)\, Y^*_{l'',m'-m}\left(\widehat{\mathbf{R}-\mathbf{R}'}\right) \tag{14}$$

is the KKR structure matrix, which is Hermitian. The summation runs over $l'' = \left|l'-l\right|,\ \left|l'-l\right|+2,\ ...,\ l'+l$, and $i^{-l+l'-l''}$ is real because $C_{LL'L''} \equiv \int Y_L(\hat{r}) Y^*_{L'}(\hat{r}) Y_{L''}(\hat{r}) d\hat{r}$. The on-site elements of $B\left(\kappa\right)$ vanish. In this way, we obtain the most important result:

$$aS^a\left(\varepsilon\right) - aD\left\{j\left(\kappa a\right)\right\} = \frac{1}{j\left(\kappa a\right)}\left[B\left(\kappa\right) + \kappa\cot\alpha\left(\kappa\right)\right]^{-1}\frac{1}{j\left(\kappa a\right)}, \tag{15}$$

where $a$, $j\left(\kappa a\right)$, $D\left\{j\left(\kappa a\right)\right\}$, and $\cot\alpha\left(\kappa\right)$ are *diagonal* matrices with elements $a_{RL}$, $j_l\left(\kappa a_{RL}\right)$, $D\left\{j_l\left(\kappa a_{RL}\right)\right\} = \kappa a_{RL} j'\left(\kappa a\right)/j\left(\kappa a\right)$, and $\cot\alpha_{RL}\left(\kappa\right)$. The quantity $a_{R'L'} S^a_{R'L',RL}\left(\varepsilon\right)$, which is $a^2_{R'L'}$ times the $L'$-component of the radial derivative of $\psi^a_{RL}\left(\varepsilon\right)$

at the $a_{R'L'}$-sphere, form the elements of a matrix which is *Hermitian.* This matrix, we call the structure matrix.

For the channels to be treated as strongly scattering with the set $\{\psi^a\}$, we take $\alpha_{RL}(\kappa)$ to be the hard-sphere phase shifts (3), and for those to be treated as weakly scattering and, thus to be *downfolded* into the SSWs, we take $\alpha_{RL}(\kappa)$ to be the real phase shifts, $\eta_{Rl}(\kappa)$. The non-scattering channels do not enter the screening calculation (15), since they neither scatter the bare, nor the screened set. The strongly- and weakly-scattering channels thus contribute to the size of the matrix to be inverted and the strongly-scattering channels are the only ones which will eventually enter the equations for solving Schrödinger's equation.

Instead of expressing (14) and (15) in terms of the usual spherical Bessel and Neumann functions, one could of course have divided the factors $\kappa^l$ and $\kappa^{-l-1}$ out on the right-hand side of (15), or used $\psi_l^0(\varepsilon, r)$ instead of $n_l(\kappa r)$, etc.. The only difference between the last equation of (2) and equation (15), is that the Hermitian matrix $aS^\alpha(\varepsilon)$ is normalized in such a way as to make its energy dependence as small as possible, and in such a way as to give $S^\alpha(\varepsilon)$ a geometrical interpretation, namely as the dimensionless slope matrix. Specifically,

$$\kappa^{-1}\tan\alpha\left[B^\alpha(\kappa)-\kappa\cot\alpha(\kappa)\right]\kappa^{-1}\tan\alpha=-j(\kappa a)\,a\left[S^a(\varepsilon)-D\left\{j(\kappa a)\right\}\right]j(\kappa a),$$

so that the screened structure matrices $B^\alpha(\kappa)$ and $aS^a(\varepsilon)$ differ because functions of energy have been subtracted from the diagonal elements, and because the rows and columns have been rescaled with such functions. If we form:

$$\mathsf{S}^{\mathsf{a}}_{R'L',RL}(\varepsilon)\equiv-2\left(w/a_{R'L'}\right)^{l'}\left[S^a_{R'L',RL}(\varepsilon)+(l+1)\,\delta_{R'L',RL}\right]\left(w/a_{RL}\right)^{l+1}, \tag{16}$$

then $\mathsf{S}^{\mathsf{a}}(0)$ is the conventional ($\kappa$=0) LMTO structure matrix for the screening constants

$$\mathsf{a}_{RL}(0)=\left[2(2l+1)\right]^{-1}\left(a_{RL}/w\right)^{2l+1},$$

and $\dot{\mathsf{S}}^{\mathsf{a}}(0)$ is its first energy derivative for some $\dot{\mathsf{a}}(0)$. For LMTO users who have developed a feeling for the sizes of the conventional structure constants and do not care about the new interpretation in terms of logarithmic derivatives, it is of course possible to use the new method in the conventional "gauge" (16). In that case, one *must* substitute the old potential functions, $\mathsf{P}^{\mathsf{a}}_{RL}(\varepsilon)$, by $-2\left(w/a_{RL}\right)^{2l+1}\left[D\left\{\varphi_{Rl}(\varepsilon,a_{RL})\right\}+l+1\right]$, with $D\left\{\varphi_{Rl}(\varepsilon,a_{RL})\right\}$ evaluated as explained in the following section.[3]

Whereas the slope matrix specifies the normal gradients on the hard spheres of all functions in the SSW set, its first energy derivative, $\dot{S}^a_{RL,R'L'}(\varepsilon)$, specifies the normal gradients of the first-energy derivative functions, $\dot{\psi}^a_{RL}(\varepsilon)$, as illustrated at the bottom of Fig. 2. Since the hard spheres are independent of energy, the energy-derivative functions will vanish at *all* hard spheres, including their own. The first energy derivative of the structure matrix in addition gives the *overlap matrix* of the SSW set: $\langle\psi^a_{RL}(\varepsilon)\,|\psi^a_{R'L'}(\varepsilon)\rangle = a\dot{S}^a_{RL,R'L'}(\varepsilon)$. This equation follows from the more general one: $\langle\psi^a_{RL}(\varepsilon)\,|\psi^a_{R'L'}(\varepsilon')\rangle = a_{RL}\left[S^a_{RL,R'L'}(\varepsilon)-S^a_{RL,R'L'}(\varepsilon')\right]/(\varepsilon-\varepsilon')$, which may be derived by use of Green's second

theorem.[3] Here, the strongly scattering radial components have been truncated inside the corresponding hard spheres as illustrated in Fig.s 1 and 2, while the remaining, regular components extend to the centers of the spheres.

Considered as functions of $\varepsilon$, the eigenvalues of the structure matrix $aS^a(\varepsilon)$ have poles when $\varepsilon$ coincides with an energy eigenvalue, $\varepsilon_i^a$, of the hard-sphere solid. For practical purposes, the $a$-radii can be chosen in such a way that the energies of interest to us are well below the bottom of the hard-sphere continuum, and below any localized state of the hard-sphere solid. For such energies, the eigenvalues of $aS^a(\varepsilon)$ are analytical functions of $\varepsilon$. This latter point is demonstrated in the left-hand side of Fig. 3, which also demonstrates that the energy dependence is weak over the $\pm 10$ eV region considered. The right-hand side shows that, for low energies or close sphere packings, the slope matrix decays by an order of magnitude per shell of neighbors. For a monotonically decaying SSW we expect, as illustrated in Fig. 2, a negative slope at its own hard sphere and positive slopes at the neighboring spheres. This is also the behavior found in Fig. 3, at least throughout the first three shells.

In conclusion, the slope matrix generated by inversion of the non singular matrix (15) contains all the information we shall need about the SSW set.

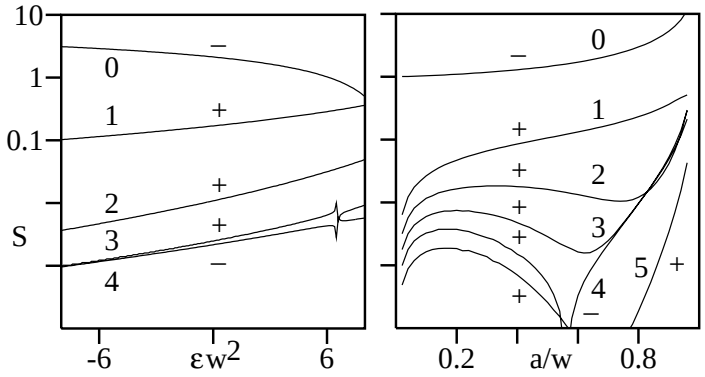


Figure 3: The $ss\sigma$-element, $S^a_{R00,000}(\varepsilon)$, of the slope matrix for the fcc structure and $R$ in the 0'th to 4'th or 5'th shell. *Left:* $a_{spd}$=0.7$w$. Here $w$ is the Wigner-Seitz radius, which in the fcc structure is 10% larger than that of touching spheres. $\varepsilon w^2$=6.05 corresponds to the lowest free-electron energy at the $X$ point. *Right:* $\varepsilon$=0 and $a_{spd}$=$a$. The $S$-scale is logarithmic and the channels with $l > 2$ were taken as non-scattering. The number of atoms in the 0'th–5'th shell are respectively 1, 12, 6, 24, 12, and 24. The calculation was performed by matrix inversion, Eq. (15), in real space for a 79-site cluster. For positive energies it was necessary to prevent resonances at the surface of this cluster by enclosing it in a concave sphere simulating the boundary condition $\psi$=0 on the spheres outside the cluster and carrying max($l$)=8. The artefact at $\varepsilon w^2 \approx$6.3 is a surviving resonance. The results are accurate only when the SSW is well localized within the cluster. For cases where the SSW's are not localized within a cluster of affordable size, we must assume crystalline boundary conditions and Bloch-sum $B(\kappa)$ with the Ewald technique before performing the screening inversion (15) for each **k**-point. With largely overlapping MT spheres, the energies of occupied states are always negative.

We now come to consider Schrödinger's equation, $[-\Delta + V(\mathbf{r}) - \varepsilon_i]\,\Psi_i(\mathbf{r}) = 0$, for a MT potential and begin by showing that with our screened spherical waves it is a rather simple matter to formulate the matching problem for the solutions, $\Psi_i(\mathbf{r})$, algebraically:

First we integrate the radial Schrödinger equation for each *strongly* scattering channel *outwards* from the origin to the MT radius $s_R$ in the potential well $v_R(r_R)$, and then *inwards* in *zero* potential (the MT zero) from $s_R$ to the hard-sphere radius $a_{RL}$. The outwards integration yields the radial partial wave $\phi_{Rl}(\varepsilon, r_R)$, and the subsequent inwards integration yields the radial partial wave "as seen from free space" $\varphi_{Rl}(\varepsilon, r_R)$, with radial logarithmic derivative $D\{\varphi_{Rl}(\varepsilon, a_{RL})\}$ at the hard sphere. These two waves match continuously and differentiably at $s_R$ and they may be seen in the left-hand side of Fig. 4, after multiplication by $Y_L(\hat{\mathbf{r}}_R)$. Let us assume that $\phi^a_{RL}(\varepsilon, r_R)$ and $\varphi^a_{RL}(\varepsilon, r_R)$ have been normalized in such a way that $\varphi^a_{RL}(\varepsilon, a_{RL}) \equiv 1$ at the hard sphere; this is what the superscript $a$ here indicates. In the case where the hard spheres have been chosen to depend on $m$, radial functions of the same $Rl$ may have different normalizations, hence the subscript $L$ rather than $l$. With this normalization, the free partial wave matches continuously, but with the kink $S^a_{RL,RL}(\varepsilon) - D\{\varphi_{Rl}(\varepsilon, a_{RL})\}$, to the $RL$-projection, $\psi^a_{R'L',RL}(\varepsilon, r_{R'})$, of the corresponding SSW, $\psi^a_{RL}(\varepsilon, \mathbf{r}_R)$. Let us furthermore truncate $\phi^a_{RL}(\varepsilon, r_R)$ and $\varphi^a_{RL}(\varepsilon, r_R)$ outside the MT sphere $(0|s_R)$ and, like the SSW, let us truncate $\varphi^a_{RL}(\varepsilon, r_R)$ also inside the $a_{RL}$-sphere. The function $[\phi^a_{RL}(\varepsilon, r_R) - \varphi^a_{RL}(\varepsilon, r_R)]\,Y_L(\hat{\mathbf{r}}_R)$ thus equals the proper partial wave inside the hard sphere, where it jumps by $-Y_L(\hat{\mathbf{r}}_R)$, and it vanishes quadratically at the MT sphere with a prefactor proportional to the MT discontinuity $v_R(s_R)$. To this function we now add the corresponding SSW thus obtaining the *kinked partial wave* (KPW):

$$\Phi^a_{RL}(\varepsilon, \mathbf{r}_R) \equiv [\phi^a_{RL}(\varepsilon, r_R) - \varphi^a_{RL}(\varepsilon, r_R)]\,Y_L(\hat{\mathbf{r}}_R) + \psi^a_{RL}(\varepsilon, \mathbf{r}_R), \tag{17}$$

which is also shown in Fig. 4. This function is everywhere continuous, but has kinks of size $S^a_{R'L',RL}(\varepsilon) - D\{\varphi_{Rl}(\varepsilon, a_{RL})\}\,\delta_{R'L',RL}$ at the hard $a_{R'L'}$-spheres.

At such a sphere, the kink of the *linear combination* of KPWs, $\sum_{RL}\Phi^a_{RL}(\varepsilon, \mathbf{r}_R)\,c^a_{RL}(\varepsilon)$, is therefore $\sum_{RL}\left[S^a_{R'L',RL}(\varepsilon) - D\{\varphi_{Rl}(\varepsilon, a_{RL})\}\,\delta_{R'L',RL}\right]\,c^a_{RL}(\varepsilon)$. If we can now find an energy, $\varepsilon_i$, and coefficients, $c^a_{RL,i}$, such that

$$\sum_{RL}\left[S^a_{R'L',RL}(\varepsilon_i) - D\{\varphi_{Rl}(\varepsilon_i, a_{RL})\}\,\delta_{R'L',RL}\right]\,c^a_{RL,i} = 0 \qquad \text{for all } R'L', \tag{18}$$

then the corresponding linear combination is *smooth* and therefore solves Schrödinger's equation with $\varepsilon_i$ as an energy eigenvalue.

The statement that the *"kink-cancellation condition"* (18) leads to a solution of Schrödinger's equation is exact only for a non-overlapping MT potential. Before continuing to the case of overlapping potentials, let us scrutinize our proof a little closer. Each KPW is constructed to be a solution of Schrödinger's equation at energy $\varepsilon$, except in all shells between concentric MT- and hard spheres, and except for the kinks at the hard spheres. In the case

where we choose all concentric hard spheres to coincide with the MT sphere ($a_{RL}$=$s_R$), all shells vanish so the statement is obviously true. That it holds also when the hard spheres are different from the concentric MT sphere, follows from the fact that for a linear combination with all kinks cancelled, each $\varphi^a_{R'L'}(\varepsilon_i, r_{R'})\, c^a_{R'L',i}$ matches the $R'L'$-projection, $\sum_{RL} \psi^a_{R'L',RL}(\varepsilon_i, r_{R'})\, c^a_{RL,i}$, of the linear combination of SSWs, $\sum_{RL} \psi^a_{RL}(\varepsilon_i, \mathbf{r}_R)\, c^a_{RL,i}$, at $a_{R'L'}$ in value *and in slope,* and since both radial functions are solutions of the *same* second-order differential equation, namely the $l'$'th radial wave equation, they must be *identical.* As a consequence,

$$\varphi^a_{R'L'}(\varepsilon_i, r_{R'})\, c^a_{R'L',i} - \sum_{RL} \psi^a_{R'L',RL}(\varepsilon_i, r_{R'})\, c^a_{RL,i} = 0, \text{ for } 0 \le r_{R'} \le s_{R'} \text{ and } R'L' \in \text{strong scat.} \tag{19}$$

Inside the $s_{R'}$-sphere then, only terms which satisfy Schrödinger's equation remain, namely $Y_{L'}(\hat{\mathbf{r}}_{R'}) \sum_{RL} \phi^a_{R'L',RL}(\varepsilon_i, r_{R'})\, c^a_{RL,i}$ and the weakly- and non-scattering channels of $\sum_{RL} \psi^a_{RL}(\varepsilon_i, \mathbf{r}_R)\, c^a_{RL,i}$.

The KPW is defined in (17) as the SSW plus the central, pure angular-momentum contribution $\phi - \varphi$, which vanishes quadratically at the MT sphere. In analogy with Slater's augmented plane wave (APW), the KPW might have been named an augmented screened spherical wave. This analogy is only complete though when all hard spheres coincide with their concentric MT sphere.

Next we consider the case of MT *overlap.* Suppose that we have solved the kink-cancellation equations (18) with logarithmic derivatives calculated for potential wells which overlap. To what extent is the resulting *smooth* function, $\Psi_i(\mathbf{r}) \equiv \sum_{RL} \Phi_{RL}(\varepsilon_i, \mathbf{r}_R)\, c_{RL,i}$, a solution of Schrödinger's equation for the superposition of these overlapping wells? The situation is sketched in Fig. 5 and the answer is, that the smooth superposition of KPWs solves Schrödinger's equation to leading (first) order in the potential overlap.

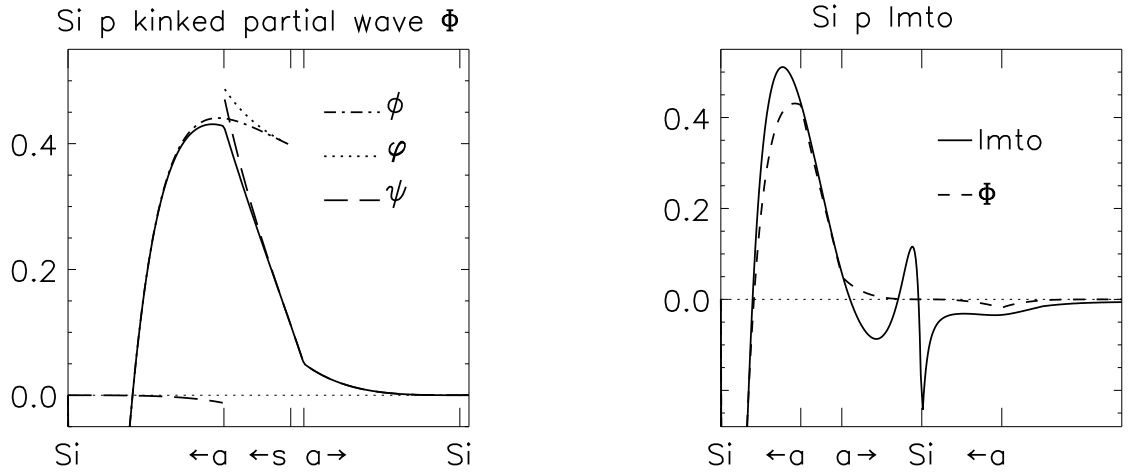


Figure 4: Kinked partial wave (KPW), $|\Phi\rangle \equiv |\phi\rangle - |\varphi\rangle + |\psi\rangle$, and LMTO, $|\chi\rangle \equiv |\Phi\rangle - \left|\dot{\Phi}\right\rangle \dot{K}^{-1} K$, for Si $p_{x+y+z}$ plotted along the [111]-direction towards a nearest neighbor in the diamond structure. Note the change of length scale between the left and right panels.

Since we have only considered the strongly-scattering channels in this one-dimensional figure, let us now be a bit more careful. Using the definition (17) and the following definitions: $\psi_i(\mathbf{r}) \equiv \sum_{RL} \psi^a_{RL}(\varepsilon_i, \mathbf{r}_R)\, c_{RL,i}$, $\phi^a_R(\varepsilon_i, \mathbf{r}_R) \equiv \sum_L \phi^a_{RL}(\varepsilon_i, \mathbf{r}_R)\, c_{RL,i}$, and similarly for $\varphi^a_R(\varepsilon_i, \mathbf{r}_R)$, we obtain:

$$\left[-\Delta + \sum_R v_R(r_R) - \varepsilon_i\right] \Psi_i(\mathbf{r}) = \sum_{R'} v_{R'}(r_{R'}) \sum_{R \neq R'} \left[\phi^a_R(\varepsilon_i, \mathbf{r}_R) - \varphi^a_R(\varepsilon_i, \mathbf{r}_R)\right]$$

$$+ \sum_R \left[-\Delta + v_R(r_R) - \varepsilon_i\right] \left[\phi^a_R(\varepsilon_i, \mathbf{r}_R) - \varphi^a_R(\varepsilon_i, \mathbf{r}_R)\right] + \left[-\Delta + \sum_R v_R(r_R) - \varepsilon_i\right] \psi_i(\mathbf{r}) =$$

$$\sum_{R'} v_{R'}(r_{R'}) \sum_{R \neq R'} \left[\phi^a_R(\varepsilon_i, \mathbf{r}_R) - \varphi^a_R(\varepsilon_i, \mathbf{r}_R)\right] - \sum_R v_R(r_R) \left[\varphi^a_R(\varepsilon_i, \mathbf{r}_R) - \psi_i(\mathbf{r})\right] - \left[\Delta + \varepsilon_i\right] \psi_i(\mathbf{r})$$

$$= \sum_{R'} v_{R'}(r_{R'}) \sum_{R \neq R'} \left[\phi^a_R(\varepsilon_i, \mathbf{r}_R) - \varphi^a_R(\varepsilon_i, \mathbf{r}_R)\right] \tag{20}$$

$$= \frac{1}{2} \sum_{R'} v_{R'}(r_{R'}) \sum_{R \neq R'} v_R(s_R) \left[(s_R - r_R)^2 + o\left((s_R - r_R)^2\right)\right] \phi^a_R(\varepsilon_i, \mathbf{r}_R) \tag{21}$$

$$\sim \frac{1}{2} \sum_{RR'}^{\text{pairs}} v_{R'}(r_{R'}) \left[(s_{R'} - r_{R'})^2 + (s_R - r_R)^2\right] v_R(s_R)\ \Psi_i(\mathbf{r}) \tag{22}$$

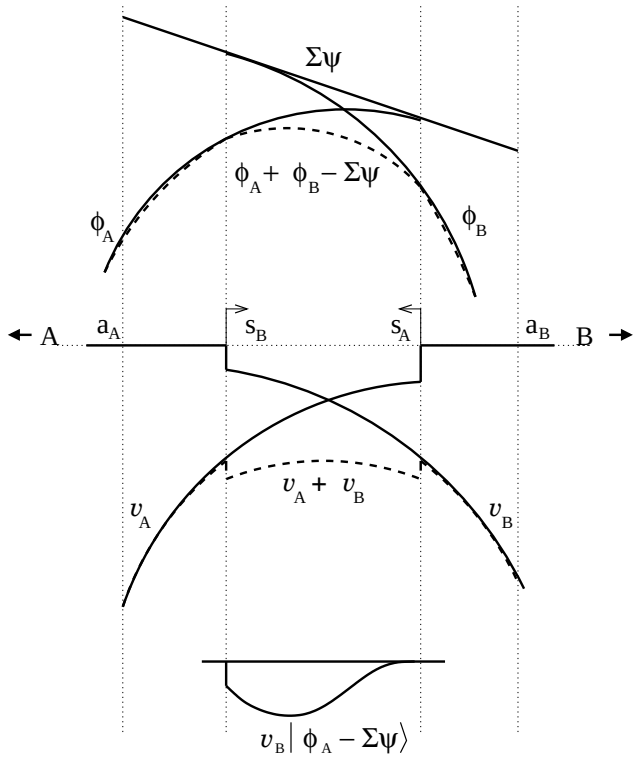


Figure 5: *Middle:* Overlapping potential wells, $v_A(r_A)$ and $v_B(r_B)$, centered at sites **A** (far left) and **B** (far right). *Top:* The solution $\phi_A \equiv \sum_L \phi_{AL}(\varepsilon_i, \mathbf{r}_A)\, c_{AL,i}$ joins smoothly onto the free solution $\varphi_A$ at $s_A$. $\varphi_A$ runs backwards to $a_A$, where the kink with the interstitial solution, $\sum\psi \equiv \sum_{RL} \psi_{RL}(\varepsilon_i, \mathbf{r}_R)\, c_{RL,i}$, is cancelled. Similarly for $\phi_B$ and $\varphi_B$. Due to kink cancellation, the resulting wave function, $\phi_A - \varphi_A + \phi_B - \varphi_B + \sum\psi$, equals $\phi_A + \phi_B - \sum\psi$ in this picture where the angular-momentum character has been suppressed. *Bottom:* The error, $[-\Delta + v_A + v_B - \varepsilon_i]\,|\phi_A + \phi_B - \sum\psi\rangle$ $= v_B\,|\phi_A\rangle + v_A\,|\phi_B\rangle - (v_A + v_A)\,|\sum\psi\rangle$ $= \quad v_B\,|\phi_A - \sum\psi\rangle \ + \ v_A\,|\phi_B - \sum\psi\rangle$ , consists of two terms each of which, *e.g.* the first, is the product of $v_B(r_B)$ and $\phi_A - \sum\psi$, which vanishes like $v_A(s_A)\,(s_A - r_A)^2\,\phi_A(s_A)$ at the $s_A$ boundary. Hence the error is of *second* order in the potential overlap.

Here, we have first of all made use of the fact that $\Psi_i(\mathbf{r})$ is smooth so that we can apply the $\Delta$-operator to its individual, kinked or discontinuous parts without keeping track of all the resulting diverging terms, because they will cancel in the end. In obtaining the 3rd line, we have used that $\phi$ solves Schrödinger's equation for its own well. Eq. (20) has then been obtained by use of Eq. (19) for the strongly-scattering partial waves, plus the fact that the weakly- and non-scattering channels ($\Lambda$) of $\psi_i(\mathbf{r})$ solve Schrödinger's equation, *i.e.* that $[\Delta+\varepsilon]\,\psi^a_{RL}(\varepsilon,\mathbf{r}_R) = \sum_{R'} v_{R'}(r_{R'}) \sum_\Lambda \psi^a_{R'\Lambda,RL}(\varepsilon,\mathbf{r}_R)$.

Returning to the strongly-scattering channels of $\sum_R v_R(r_R)\,[\varphi^a_R(\varepsilon_i,\mathbf{r}_R) - \psi_i(\mathbf{r})]$, if the overlap is so large that the $s_R$-sphere overlaps a *neighboring* $a_{R'L'}$-sphere, then it is simplest to imagine that we have *not* truncated the SSWs inside their hard spheres, because otherwise the cancellation (19) would not take place inside the $s_R$-$a_{R'L'}$ overlap. For consistency then, we should not truncate the free partial waves $\varphi$ inside their own hard sphere either. The resulting divergencies at the sites, of the SSWs and of the free partial waves, of course cancel for the *smooth* linear combinations. This undoing of the truncation inside the hard spheres is not necessary, but it simplifies the bookkeeping.

The result (20) is then in agreement with what we found in Fig. 5, that the error is a function which vanishes outside the regions of overlap and that inside such a region, it is the product of a function, $v_{R'}(r_{R'})$, which vanishes with a small discontinuity at one of the MT spheres and a function, $\phi_R - \varphi_R$, which vanishes quadratically at the surface of the other MT sphere, with a prefactor proportional to the discontinuity of that MT potential. Remember that the radial part of $\varphi_R$ is supposed to continue to the origin. The result, which is given in (21), may be obtained from the radial Schrödinger and wave equations. Finally, in expression (22) we have kept only the term of leading order and have used that $\phi^a_R(\varepsilon_i,\mathbf{r}_R) \approx \Psi_i(\mathbf{r})$ in the region picked out by the other factors. Hence, *the error of the wave function is of second order in the potential overlap.*

The error of the one-electron energy may be obtained by first order perturbation theory as: $\Delta\varepsilon_i \equiv \varepsilon_i - \varepsilon_i^{\text{true}} \approx -\langle\Psi_i|-\Delta+\sum_R v_R(r_R)-\varepsilon_i|\Psi_i\rangle$ and, to leading order, we find from Eq. (22) that the error of the band-structure energy is[2]

$$\sum_i^{\text{occ}} \Delta\varepsilon_i \;\sim\; -\frac{\pi}{24}\sum_{RR'}^{\text{pairs}} |\mathbf{R}-\mathbf{R}'|^5\,\omega^4_{RR'}\,v_R(s_R)\,v_{R'}(s_{R'})\,\rho\left(\frac{\mathbf{R}+\mathbf{R}'}{2}\right), \tag{23}$$

$$\text{where}\quad \omega_{RR'} \;\equiv\; \frac{s_R+s_{R'}}{|\mathbf{R}-\mathbf{R}'|}-1 \quad \text{is the radial overlap.} \tag{24}$$

In a last section we shall demonstrate how this works for the third-generation LMTO method. Finally, it may be noted that the appearance of the KPW in Fig. 4 is hardly influenced by the MT overlap. This figure in fact applies to an overlap of $\omega$=14%.

Like the slope matrix, the kink matrix is not Hermitian, but the matrix

$$K^a_{R'L',RL}(\varepsilon) \equiv a_{R'L'}\left[S^a_{R'L',RL}(\varepsilon) - D\{\varphi_{Rl}(\varepsilon,a_{RL})\}\,\delta_{R'L',RL}\right] \tag{25}$$

is.[21] This matrix is the renormalized screened KKR matrix. If we multiply each of the kink-cancellation equations (19) with the corresponding hard sphere radius, $a_{R'L'}$, these equations

take the form: $\sum K\mathbf{c}=\mathbf{0}$ and, hence, they are the screened KKR equations. Just as the first energy derivative of the structure matrix is the overlap matrix for the set of SSWs, so the first energy derivative of the KKR matrix is the overlap matrix for the set of KPWs.[22] In fact, one may show that the KKR matrix itself is the energy minus the MT Hamiltonian in the basis of the KPWs with the same energy, that is,

$$K^a_{R'L',RL}\left(\varepsilon\right)=\left\langle \Phi^a_{R'L'}\left(\varepsilon\right)\left|\varepsilon-\left(-\Delta+V\right)\right|\Phi^a_{RL}\left(\varepsilon\right)\right\rangle . \tag{26}$$

For Green-function and CPA calculations it has been very important that the transformation (7) of the resolvent, $\left[\mathsf{P}\left(z\right)-\mathsf{S}\right]^{-1}$, from one representation to another is merely a scaling rather than a matrix operation. This turns out to hold also in the new formalism, and it means that such calculations may now be performed with more realistic potentials and including downfolding. The result is:

$$K^b\left(z\right)^{-1}=a^{-1}g^a\left(z,b\right)\varphi^a\left(z,b\right)+\varphi^a\left(z,b\right)\,K^a\left(z\right)^{-1}\,\varphi^a\left(z,b\right) \tag{27}$$

and has been obtained by use of the completeness relation (12), the one-centre expansion (13), and the following Wronskian relations:[3]

$$ag^b\left(a\right)=-bg^a\left(b\right),\quad af^b\left(a\right)=b^2g^a\left(b\right)',\quad a^2g^b\left(a\right)'=bf^a\left(b\right),\quad a^2f^b\left(a\right)'=-b^2f^a\left(b\right)',$$

where the common energy argument, $z$, has been dropped.

LOW-ENERGY, FEW-ORBITAL, TB HAMILTONIANS; HTSCs

If the energy dependence of the renormalized screened KKR matrix is linearized around some chosen energy $\varepsilon_\nu$,

$$K^a\left(\varepsilon\right)\approx\,K^a+\left(\varepsilon-\varepsilon_\nu\right)\dot{K}^a=-\left\langle\Phi^a\left|-\Delta+V\right|\Phi^a\right\rangle+\varepsilon\left\langle\Phi^a\mid\Phi^a\right\rangle, \tag{28}$$

then the KKR equations (18) have the form of an algebraic eigenvalue problem. In (28) and in the following, omission of an energy argument $\varepsilon$ means that the function is evaluated *at* $\varepsilon_\nu$. The basis set which, by use of the Raleigh-Ritz variational principle for the MT Hamiltonian, gives rise to this problem turns out to be the KPW-set at the fixed energy, $\varepsilon_\nu$. This follows from Eq. (26) and is expressed in the second part of Eq. (28). Since the off-diagonal elements of the overlap matrix, $\dot{K}$, only influence the energy eigenvalues to order $\left(\varepsilon_i-\varepsilon_\nu\right)^2$, we may even neglect the non-orthogonality of the KPWs and, for a crystal, obtain the correct Fermi surface, $\varepsilon_i\left(\mathbf{k}\right)=\varepsilon_F\equiv\varepsilon_\nu$, and the correct group velocities, $\partial\varepsilon_i\left(\mathbf{k}\right)/\partial\mathbf{k}\big|_{\varepsilon_F}$, by diagonalization of a first-order Hamiltonian whose matrix elements are simply: $-\tilde{K}^a_{RL,R'L'}\equiv-K^a_{RL,R'L'}\Big/\sqrt{\dot{K}^a_{RL,RL}\dot{K}^a_{R'L',R'L'}}$. This Hamiltonian is completely analogous to $h^{\mathsf{a}}$ in the ASA, but $-\tilde{K}^a$ implicitly contains the integrals over the interstitial region and the downfolded channels, and it works to leading order in the overlap of the potential wells. The range of $-K_{RL,R'L'}$ in $R$-space, and the size of the energy window inside which the linear approximation holds, depends on the screening. Crudely speaking, the more

strongly-scattering channels included, and the larger their hard spheres chosen without being touching, the shorter is the range of the hopping, and the wider is the energy window.

$-\tilde{K}^a$ can be used as the low-energy, few-orbital, single-particle part of correlated Hubbard-type Hamiltonians, as we shall now demonstrate for a generic high-temperature superconductor (HTSC). We have in the past[23] been able to derive such a Hamiltonian for $YBa_2Cu_3O_7$ using the second-generation LMTO package.[16] That procedure, however, required a lot of hand-work and much insight, and has proved cumbersome to use in general. The new procedure is far more automatic and accurate,[24] and has already proved successful for the ladder compounds.[25]

The basic structural element of all HTSCs is a $CuO_2$ layer, which is a quadratic lattice with copper at the corners and oxygen halfway between all copper nearest neighbors. In the left-hand side of Fig. 6, the copper sites are those which carry either a $d_{x^2-y^2}$ or an $s$ orbital, and the oxygen sites are those which carry a $p_x$ or a $p_y$ orbital. Different HTSC materials have different stackings of the $CuO_2$ layers with various "insulating" and/or "doping" layers between them. Nevertheless, the calculated LDA band structures near what is believed to be the Fermi level of optimally doped HTSCs are very similar, and similar to that calculated for the simplest possible such material; dimpled $CaCuO_2$. In this compound, the $CuO_2$ layers are stacked in the $z$-direction and are separated by calcium, which sits in the hollow between the eight coppers of the two neighboring layers. The oxygens in the Cu rows running in the $x$- ($y$-)direction are dimpled out of the plane by $+(-)$ 7 degrees. The right-hand side of Fig. 6 shows a central $CuO_2$ layer seen from the side, with a $d_{x^2-y^2}$ and an $s$ orbital on the copper sites and a $p_z$ orbital on the oxygen site. On the $CuO_2$ layer above is shown a Cu $s$ orbital and on the $CuO_2$ layer below, an O $p_z$ orbital. Dimpled $CaCuO_2$ is a calculated structure,[26] a theorists dream which hardly exists in this simple form in nature. Its LDA energy bands, which we shall now consider, are nevertheless very similar to those calculated[23] for $YBa_2Cu_3O_7$, one of the only known *stoichiometric* optimally doped HTSCs.

At the Fermi level there is only *one* band per $CuO_2$ layer, and this is the anti-bonding $pd\sigma$ band formed from the O $p_x$ – Cu $d_{x^2-y^2}$ – O $p_y$ orbitals. This band is at the top of the 10 eV broad O $p$ – Cu $d$ complex consisting of 16 bands, the upper (anti- and non-bonding) part of which may be seen in Fig. 7 (a). According to the LDA and the so-called Van Hove scenario of HTSC, the Fermi level (zero in the figure) for the optimally doped compounds is very close to the saddle-point of the conduction band at $(ak_x, ak_y) = (\pi, 0)$. Hybridization with the Cu $s$ band, which is 5 eV above, has pushed this saddle-point of the anti-bonding $pd\sigma$ band down in energy, to a point where it just "straddles off" the top of the anti-bonding $pd\pi$ bands. This makes the structure susceptible to out-of-row movements of oxygen, because this will mix $\sigma$ and $\pi$ bands. In particular the stable structures of $CaCuO_2$ and $YBa_2Cu_3O_7$ have oxygen dimpled seven degrees out of the layer, and this mixes O $p_z$ character into the conduction band in such a way that its saddle-point at $(\pi, 0)$ becomes "extended" that is, the dispersion towards $(0, 0)$ becomes proportional to $k^4$, *i.e. flat,* while in the perpendicular direction, towards $(\pi, \pi)$, it remains $k^2$. The mixing pushes the corresponding $pd\pi$ band

down in energy by about half an eV and leaves the top of other $pd\pi$ bands about an eV below the Fermi level. We thus see that the orbital character of the conduction band, which is the only one we wish to describe, is quite mixed.

The converged LDA bands are showed in panel (a) of Fig. 7. For comparison, panels (b)-(d) show the bands calculated with various "minimal" LMTO sets, specifically, with only the six O $p$ orbitals (b), with only the Cu $d_{x^2-y^2}$ orbital (c), and with the six O $p$ orbitals plus the Cu $s$ and $d_{x^2-y^2}$ orbitals (d). These four calculations all employ the full 3rd-generation LMTO formalism, to be described in the following section, in which the Hamiltonian and overlap matrices, (31) and (30), are given in terms of $K^a(\varepsilon_F)$ and its first three energy derivatives. Panel (b) and (c) demonstrate the power of downfolding in the 3rd-generation LMTO scheme: One may for instance completely leave out the Cu $d_{x^2-y^2}$ LMTOs by attaching that partial-wave character to the tails of the neighboring O $p$ LMTOs (b), or one may completely leave out the O $p$ LMTOs, keeping per cell just the one Cu $d_{x^2-y^2}$ LMTO whose tail then incorporates the O $p$, Cu $s$, and other characters (c). As one can imagine, such *massive* downfolding leads to *long* range of the LMTOs. As an example, the Fourier transform of the conduction band shown in panel (c) is the two-centre Hamiltonian in the representation of orthogonalized Cu $d_{x^2-y^2}$ LMTOs, where the cone-like feature of the band around $(0,0)$, caused by near degeneracy of the Cu $d_{x^2-y^2}$ and O $p_x$ orbital energies, gives rise to very long range. This long-ranged, single-band Hamiltonian, we have called (the single-particle part of) the "physical" low-energy Hamiltonian.[23]

What we shall be interested in here is a "chemical" Hamiltonian, which has short range and whose TB parameters behave in a meaningful way when the structure is deformed and

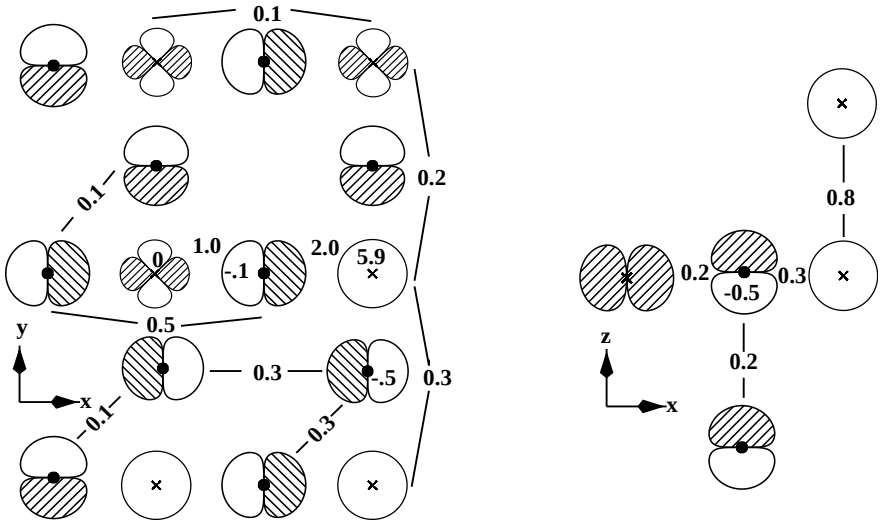


Figure 6: The eight orbitals [$O_x$ $p_{x,y,z}$, $O_y$ $p_{x,y,z}$, Cu $s$, and Cu $d_{x^2-y^2}$] and the values of their energies (with respect to $\varepsilon_{x^2-y^2}$) and two-centre hopping integrals (eV). These values were obtained as the matrix elements of $-\tilde{K}^a(\varepsilon_F)$ with all channels other than the eight downfolded.

when we proceed to similar materials. Which orbitals such a chemical Hamiltonian should contain is then dictated by the range of the corresponding $K^a(\varepsilon)$ matrix. If we imagine a Taylor series like (28), it is conceivable that the higher energy-derivative matrices have longer range. We therefore expect to obtain the shortest range when the energy region of interest is so small that we only need $\tilde{K}^a_{RL,R'L'}(\varepsilon_F)$ as defined above. For dimpled $CaCuO_2$, the chemical basis set turns out to be the one used to generate the bands shown in panel (d). For the same eight orbitals, we show in panel (e) the bands calculated by diagonalization of the effective two-center Hamiltonian $-\tilde{K}(\varepsilon_F)$. We see that this approximation conserves the shape of the conduction band in the relevant range of energy. All computations illustrated

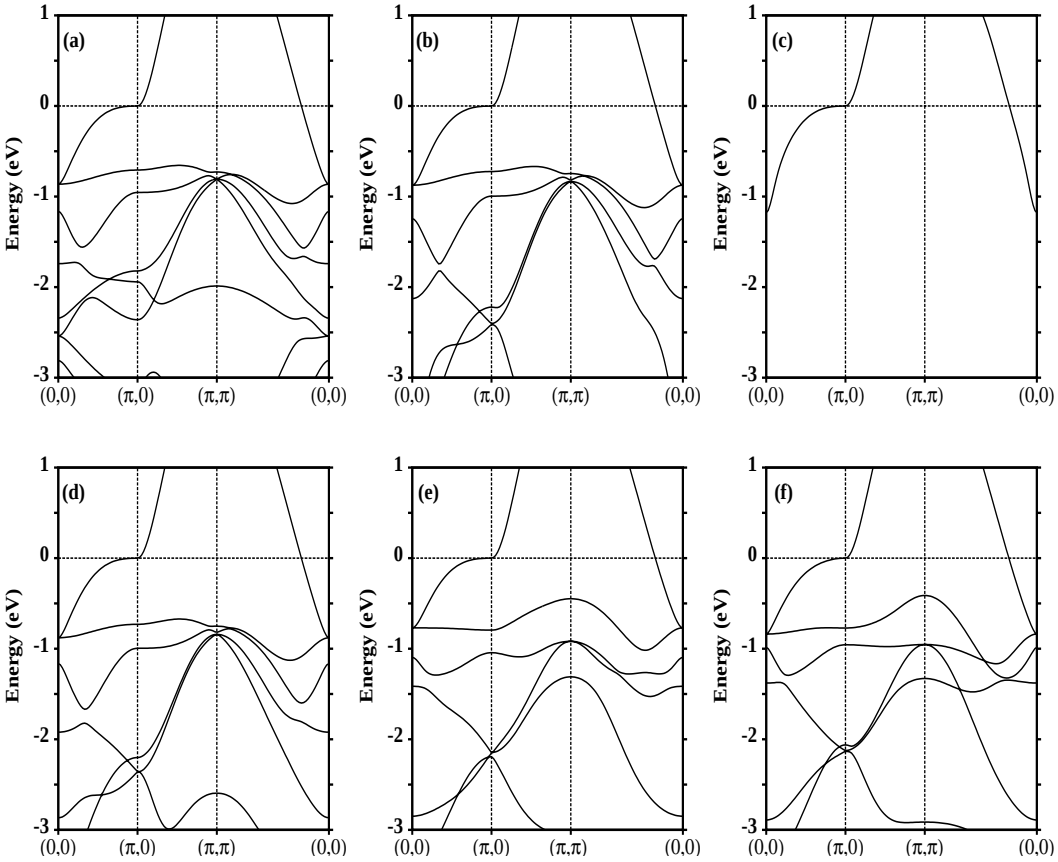


Figure 7: LDA energy bands for dimpled $CaCuO_2$ calculated with the 3rd-generation LMTO method using a converged LMTO basis, *(a),* and five different simplifications, *(b)-(f).* Reciprocal-space distances are in units of the reciprocal of the Cu-Cu distance. $\varepsilon_F = 0 = \varepsilon_\nu$. *(a):* All but the Cu $s$ and $d$, and the O $p$ orbitals downfolded. $a_{\mathrm{Cu}\,s}$=$a_{\mathrm{Cu}\,d}$=0.87$t$, $a_{\mathrm{O}\,p}$=0.75$t$, where the touching-sphere radius, $t$, is 1/4 the Cu-Cu distance. *(b):* All but the six oxygen $p$ orbitals downfolded. $a_{\mathrm{O}\,p}$=0.96$t$. *(c):* All but the single Cu $d_{x^2-y^2}$ orbital downfolded. $a_{\mathrm{Cu}\,x^2-y^2}$=0.62$t$. *(d):* All but the six oxygen orbitals, the Cu $s$, and the Cu $d_{x^2-y^2}$ orbitals downfolded. $a_{\mathrm{Cu}\,s}$=0.87$t$, $a_{\mathrm{Cu}\,x^2-y^2}$=0.68$t$, $a_{\mathrm{O}\,p}$=0.96$t$, *(e):* Like *(d),* but with $H-\varepsilon_F = -\tilde{K}$ and $O = 1$. *(f):* Like *(e),* but with $\tilde{K}$ truncated after 3rd-nearest-neighbor hoppings, that is, with the orbital energies and two-centre hopping integrals given in Fig. 6.

so far were converged in **R**-space. In fact, they were performed in **k**-space, which means that we started out using the Ewald method to compute $B(\kappa, \mathbf{k})$. When we now Fourier transform $\tilde{K}^a_{RL,R'L'}(\varepsilon_F, \mathbf{k})$, we find that the only non-negligible matrix elements are those given by the orbital energies and two-center hopping integrals in Fig. 6. Panel (f) of Fig. 7 shows the corresponding TB energy bands. This orthogonal, two-center TB Hamiltonian, is seen to reproduce the conduction band very well and to give a satisfactory description of the neighboring bands. This TB Hamiltonian, which we have generated almost automatically, could also have been calculated without the Ewald scheme, by inversion of Eq. (15) in **R**-space. A bit of trial and error is still needed in finding an optimal choice of the hard-sphere radii. The ones we used are listed in the figure caption.

LINEAR MUFFIN-TIN ORBITALS

The first-order Hamiltonian $-\tilde{K}$ does not suffice to describe the energy spectrum over the 10-20 eV range spanned by the valence and lower conduction bands of strongly bonded materials. Nor does inclusion of terms beyond the linear in the Taylor series (28) help, because this does not lead to an algebraic eigenvalue problem. What is needed, is a set of energy independent orbitals which, in contrast to the set of KPWs at a fixed energy, is complete to *linear* order in $\varepsilon - \varepsilon_\nu$.

From a set of KPWs, we first define a set of energy dependent MTOs:

$$|\chi(\varepsilon)\rangle \equiv |\Phi(\varepsilon)\rangle - \left|\dot{\Phi}\right\rangle \dot{K}^{-1} K(\varepsilon) \tag{29}$$

Here and in the following we often drop the common superscript $a$, and omission of an energy argument means that $\varepsilon{=}\varepsilon_\nu$. Moreover, we have used the notation in which $|\chi(\varepsilon)\rangle$ is a row vector with elements $|\chi_{RL}(\varepsilon)\rangle \equiv \chi_{RL}(\varepsilon, \mathbf{r}_R)$ and $K$ is a matrix. $\dot{\Phi}_{RL}(\mathbf{r}_R)$ is the first energy derivative at $\varepsilon_\nu$ of the KPW, $\Phi_{RL}(\varepsilon, \mathbf{r}_R)$, defined in (17). Since the hard spheres are kept independent of energy, the strongly-scattering channels of the energy-derivative functions $\dot{\Phi}$ vanish at *all* the hard spheres. The $\dot{\psi}$-part is sketched in the bottom half of Fig. 2 and the Si $p_{x+y+z}$ MTO at energy $\varepsilon_v$, that is the LMTO, is shown together with the corresponding KPW in the right-hand side of Fig. 4.

The superposition of $\dot{\Phi}$-functions added to the KPW in (29) is such as to make the MTO *smooth.* That this is so is seen immediately by forming the kink matrix for the MTO: $K(\varepsilon) - \dot{K}\dot{K}^{-1}K(\varepsilon) = 0$. Still, the set of MTOs remains *complete* with respect to the MT potential, because with $\varepsilon_i$ being the energy and $\mathbf{c}_i$ a corresponding solution of the KKR equations, $K(\varepsilon_i)\,\mathbf{c}_i = \mathbf{0}$, we find that the same linear combination of MTOs is: $|\chi(\varepsilon_i)\rangle\,\mathbf{c}_i = |\Phi(\varepsilon_i)\rangle\,\mathbf{c}_i = |\Psi_i\rangle$. In contrast to the KPW, the MTO is *independent of energy to linear order* because by differentiation of (29) with respect to energy and subsequent setting $\varepsilon{=}\varepsilon_\nu$ we get: $|\dot{\chi}\rangle = \left|\dot{\Phi}\right\rangle - \left|\dot{\Phi}\right\rangle \dot{K}^{-1}\dot{K} = 0$. The energy-independent set of LMTOs, $|\chi\rangle \equiv |\Phi\rangle - \left|\dot{\Phi}\right\rangle \dot{K}^{-1}K$, is therefore complete to linear order with respect to the MT Hamiltonian and therefore yields eigenvalues with errors proportional to $(\varepsilon_i - \varepsilon_\nu)^4$. For comparison the conventional single-$\kappa$ LMTO set is complete to zeroth order in the MT interstitial, albeit to first order in

the spheres, and therefore yields eigenvalue errors of order $(\varepsilon_i - \varepsilon_\nu)^2$ which originate from the interstitial. This is illustrated in Fig. 8. A price for carrying not only $\psi$, but also $\dot{\psi}$ functions, is that the new LMTO sets corresponding to different hard-sphere radii are no longer linear combinations of each other; the wave-function error, $A^a \cdot (\varepsilon_i - \varepsilon_\nu)^2$, has an $a$-dependent prefactor.

We now derive the expressions for the Hamiltonian and overlap matrices in the new LMTO basis. For the integrals in all space of KPWs and their first energy derivative functions, one obtains: $\langle \Phi \mid \Phi \rangle = \dot{K}$, $\left\langle \Phi \mid \dot{\Phi} \right\rangle = \left\langle \dot{\Phi} \mid \Phi \right\rangle = \frac{1}{2!}\ddot{K}$, and $\left\langle \dot{\Phi} \mid \dot{\Phi} \right\rangle = \frac{1}{3!}\ \dddot{K}$ . The LMTO *overlap matrix* is therefore:

$$\begin{aligned} \langle \chi \mid \chi \rangle &= \langle \Phi \mid \Phi \rangle - \left\langle \Phi \mid \dot{\Phi} \right\rangle \dot{K}^{-1} K - K \dot{K}^{-1} \left\langle \dot{\Phi} \mid \Phi \right\rangle + K \dot{K}^{-1} \left\langle \dot{\Phi} \mid \dot{\Phi} \right\rangle \dot{K}^{-1} K \\ &= \dot{K} - \frac{1}{2!}\left( \ddot{K} \dot{K}^{-1} K + K \dot{K}^{-1} \ddot{K} \right) + \frac{1}{3!} K \dot{K}^{-1}\ \dddot{K}\ \dot{K}^{-1} K. \end{aligned} \tag{30}$$

The matrix elements of the MT Hamiltonian used to generate the LMTO set may be found in a similar way. Since the LMTO is smooth there are no problems with Hermiticity like those occurring for the matrix elements between KPWs *alone*. What we mean is, that the result (26) cannot be obtained by naively taking matrix elements of an equation like: $[H - \varepsilon]\,|\Phi(\varepsilon)\rangle = 0$, where $H \equiv -\Delta + V$, or of its energy derivative: $[H - \varepsilon_\nu]\left|\dot{\Phi}\right\rangle = |\Phi\rangle$ . For matrix elements between *smooth* linear combinations of KPWs like:

$$\langle \chi \,| -\Delta + V - \varepsilon_\nu |\, \chi \rangle = \langle \Phi \,| H - \varepsilon_\nu |\, \Phi \rangle - \left\langle \Phi \,| H - \varepsilon_\nu |\, \dot{\Phi} \right\rangle \dot{K}^{-1} K - K \dot{K}^{-1} \left\langle \dot{\Phi} \,| H - \varepsilon_\nu |\, \Phi \right\rangle$$

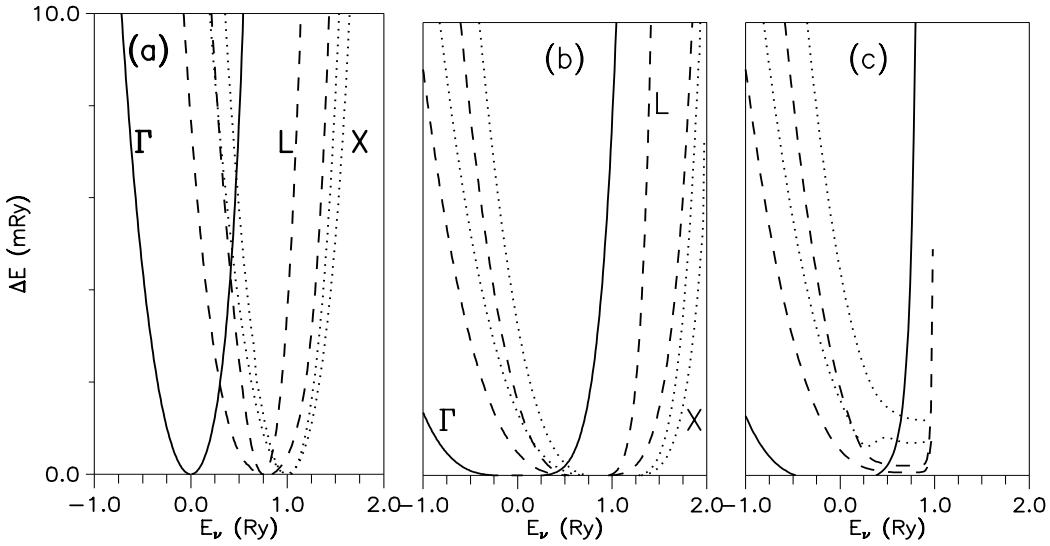


Figure 8: LMTO errors of the lowest free-electron energies (0, 0.75, and 1 Ry) at the fcc Γ, L, and X points as functions of the energy-expansion parameter $\varepsilon_\nu$. The states at L and X are doubly degenerate and split when $\varepsilon_\nu \neq \varepsilon$. *(a):* Old LMTO method with *spd*-basis and $s$=$w$. *(b):* New LMTO method, Eq.s (30) and (31), with *spd*-basis and $a_{spd}$=$0.7w$=$0.77t$; the $s$-value is irrelevant. In *(a)* and *(b)* the bare structure matrix was Bloch-summed with the Ewald technique before it was screened by matrix inversion, Eq. (15), in **k**-space. *(c):* Like *(b)* except that the inversion (15) was performed in **R**-space using a 79-site cluster enclosed in a concave sphere.

$$
\begin{aligned}
& +K\dot{K}^{-1}\left\langle \dot{\Phi}\left|H-\varepsilon_{\nu}\right|\dot{\Phi}\right\rangle \dot{K}^{-1}K \\
= & -\left\langle \Phi\left|H-\varepsilon_{\nu}\right|\dot{\Phi}\right\rangle \dot{K}^{-1}K+K\dot{K}^{-1}\left\langle \dot{\Phi}\left|H-\varepsilon_{\nu}\right|\dot{\Phi}\right\rangle \dot{K}^{-1}K \\
= & -\left\langle \Phi\mid\Phi\right\rangle \dot{K}^{-1}K+K\dot{K}^{-1}\left\langle \dot{\Phi}\mid\Phi\right\rangle \dot{K}^{-1}K \\
= & -K+\frac{1}{2!}K\dot{K}^{-1}\ddot{K}\dot{K}^{-1}K
\end{aligned}
\tag{31}
$$

such procedures are however correct when used consistently for all terms. Expression (31) thus gives the *MT Hamiltonian matrix* which, together with the overlap matrix (30), are given exclusively in terms of $K,\ \dot{K},\ \ddot{K},$ and $\dddot{K}$ . These matrices are square and labelled by the channels of the strong scatterers. We stress, that in the 3rd-generation LMTO method, downfolding takes place at the screening stage (15), where it removes the weakly-scattering channels from the structure matrix $S\left(\varepsilon\right)$ . The calculations for $CaCuO_2$ presented in Fig. 7 (a)-(d) employed this formalism and convincingly demonstrated the new downfolding.

An approximation, which goes beyond the ASA and is not based on dividing space into spheres and neglecting the remainder, consists of neglecting all off-diagonal elements in the *real*-space representation of $\dot{K},\ \ddot{K},$ and $\dddot{K}$. With this *new* ASA, we have avoided the matrix inversion, $\dot{K}^{-1}$, and the formalism contains only *one* matrix, which we may take to be the first-order two-centre Hamiltonian $-\tilde{K}$ defined in the previous section. This corresponds to renormalizing each KPW and each MTO according to: $\left|\tilde{\Phi}_{RL}\left(\varepsilon\right)\right\rangle \equiv\ \left|\Phi_{RL}\left(\varepsilon\right)\right\rangle \Big/\sqrt{\left\langle \Phi_{RL}^{2}\right\rangle}\ =\ \left|\Phi_{RL}\left(\varepsilon\right)\right\rangle \Big/\sqrt{\dot{K}_{RL,RL}}$ and $\left|\tilde{\chi}_{RL}\left(\varepsilon\right)\right\rangle \equiv\ \left|\chi_{RL}\left(\varepsilon\right)\right\rangle \Big/\sqrt{\dot{K}_{RL,RL}}$ , and the rows and columns of the KKR matrix accordingly: $\tilde{K}_{RL,R'L'}\left(\varepsilon\right)\equiv K_{RL,R'L'}\left(\varepsilon\right)\Big/\sqrt{\dot{K}_{RL,RL}\dot{K}_{R'L',R'L'}}$ . With this renormalization, and taking $\varepsilon{=}\varepsilon_{\nu}$, it is easy to see that expressions (30) and (31) reduce to the simple ASA form (8) and (9).

We can develop an exact formalism by Löwdin orthonormalizing the KPWs, instead of merely normalizing them: The overlap matrix for the renormalized KPWs is:

$$
\left\langle \tilde{\Phi}_{RL}\mid\tilde{\Phi}_{R'L'}\right\rangle =\dot{\tilde{K}}_{RL,R'L'}\equiv\delta_{RL,R'L'}+\Delta_{RL,R'L'},
\tag{32}
$$

where $\Delta$ is a Hermitian matrix with vanishing diagonal in $RL$-space. Its off-site elements ($R\neq R'$) are usually considerably smaller than unity and if we now define a Hermitian matrix: $\dot{\tilde{K}}^{-1/2}=\left(1+\Delta\right)^{-\frac{1}{2}}\equiv\ 1-\frac{1}{2}\Delta+\frac{3}{8}\Delta^{2}-...$, which is the power-series expansion in $\Delta$, then the linear combinations $\left|\bar{\Phi}\left(\varepsilon\right)\right\rangle \equiv\left|\tilde{\Phi}\left(\varepsilon\right)\right\rangle \dot{\tilde{K}}^{-1/2}$ are seen to form an orthonormal set when $\varepsilon{=}\varepsilon_{\nu}$. This is formally like in the conventional ASA. The partial waves truncated outside and normalized inside the atomic $s$-spheres become in the formalism of the 3rd generation the Löwdin orthonormalized kinked partial waves. The transformed MTO set is: $\left|\bar{\chi}\left(\varepsilon\right)\right\rangle \equiv\ \left|\tilde{\chi}_{RL}\left(\varepsilon\right)\right\rangle \dot{\tilde{K}}^{-1/2}=\ \left|\bar{\Phi}\left(\varepsilon\right)\right\rangle +\left|\dot{\bar{\Phi}}\right\rangle h\left(\varepsilon\right)$, where

$$
h\left(\varepsilon\right)\equiv-\dot{\tilde{K}}^{-1/2}\ \tilde{K}\left(\varepsilon\right)\dot{\tilde{K}}^{-1/2}=-\left(1-\frac{1}{2}\Delta+\frac{3}{8}\Delta^{2}-...\right)\tilde{K}\left(\varepsilon\right)\left(1-\frac{1}{2}\Delta+\frac{3}{8}\Delta^{2}-...\right).
\tag{33}
$$

Since this expression for the MTO set is also formally identical with an expression which, with the old definitions, was valid only in the ASA, everything else works out the same. *E.g.,* we find: $\left|\dot{\bar{\chi}}\right\rangle = \left|\dot{\bar{\Phi}}\right\rangle + \left|\bar{\Phi}\right\rangle \dot{h} = 0$, because $\dot{h} = -1$. The Hamiltonian and overlap matrices are thus given by (8) with $h \equiv h\left(\varepsilon_\nu\right)$,

$$o \equiv \left\langle \bar{\Phi} \left| \dot{\bar{\Phi}} \right.\right\rangle = \left\langle \dot{\bar{\Phi}} \right| \left. \bar{\Phi} \right\rangle = -\frac{\ddot{h}}{2!}, \text{ and } p + o^2 \equiv \frac{1}{2}\left\langle \bar{\Phi} \left| \ddot{\bar{\Phi}} \right.\right\rangle = \frac{1}{2}\left\langle \ddot{\bar{\Phi}} \right| \left. \bar{\Phi} \right\rangle = \left\langle \dot{\bar{\Phi}} \right| \left. \dot{\bar{\Phi}} \right\rangle = -\frac{\dddot{h}}{3!}. \quad (34)$$

In the 3rd-generation LMTO, $h$, $o$, and $p$ are square matrices labelled by the strongly-scattering channels. What we have accomplished is therefore to transform the new Hamiltonian and overlap matrices, (31) and (30), into the form (8), which was previously valid only in the ASA.

In this language the *new* ASA corresponds to neglecting $\Delta$ as well as the off-diagonal, real-space parts of $o$ and $p$. A better approximation is to keep $\Delta$ to first order in Eq. (33), and *then* to neglect the off-diagonal parts in the real-space representation of $o$ and $p$. In this way we still need to specify only *one* matrix, namely the first-order, two-center TB Hamiltonian, $h$, at the expense of increasing its real-space range somewhat beyond that of $-\tilde{K}$. In the full formalism we have to specify 2 matrices, the Hamiltonian and the overlap matrix or worse, the 3 matrices: $h$, $\left(\dot{h} = -1\right)$, $\ddot{h}$, and $\dddot{h}$, or even worse, the 4 matrices: $K$, $\dot{K}$, $\ddot{K}$, and $\dddot{K}$ whose real-space range increases with the number of energy derivatives taken, that is, in order of decreasing importance for the bands near $\varepsilon_\nu$.

Some of this is illustrated in Fig. 9 where we compare the LDA band structure obtained from a converged 3rd-generation LMTO calculation (full line) with results (dashed lines) obtained using various *minimal* basis sets, $sp^3$ in (a)-(c) and $sp^3d^5$ in (d), and various *truncations.* The empty-sphere *spd*- and, in (a)-(c), the Si *d*-channels were downfolded. Here panel (a) demonstrates that it is possible with merely an $sp^3$ set to obtain an accurate first-principles description of the valence *and* four lowest conduction bands, provided that we allow the set to be so long ranged that its Hamiltonian and overlap matrices, (31) and (30), extend to 12th-nearest neighbors. This basis is defined by: $a_s$=1.1$t$, $a_p$=1.0$t$, and $\varepsilon_\nu = -2$ eV. As usual, $t$ is half the nearest-neighbor distance. If an accurate $sp^3$ TB-description is needed of merely the valence band, then it is possible to limit the range of the orbitals to the extent that the Hamiltonian and overlap matrices can be truncated after the 6th-nearest neighbors. In (b) this is achieved mainly by shifting $\varepsilon_\nu$ down to the middle of the valence band. In (c) and (d) we have simplified the calculation of the Hamiltonian and overlap matrices by evaluating (33) to only first order in $\Delta$, and by neglecting the off-diagonal elements in $R$-space of $o$ and $p$. As mentioned above, this also makes it necessary to tabulate only *one* two-centre matrix, $h$. (Note that the *screened* two-centre matrices cannot be completely specified by Slater-Koster two-centre integrals like (5), because the *screened* KPWs and LMTOs do *not* have pure angular-momentum character). Comparison of the dashed lines in (b) and (c) shows that this simplification works for the valence-band structure, but that the quality of the conduction band, which was not aimed at here, has deteriorated.

So far we have not been able with our first-principles procedure to find parameters which will decrease the range of the $sp^3$ first-order two-centre Hamiltonian, $h$, below 6th-nearest neighbors. However, with an $sp^3d^5$ basis this is possible, because then also the $d$-channels can be used for screening. This is demonstrated in panel (d), where the $sp^3d^5$-set with the parameters $a_s$=$a_p$=1.0$t$, $a_d$=0.9$t$, and $\varepsilon_\nu = -6$ eV, plus the above-mentioned simplification, yields an $h$ which can be truncated after 3rd-nearest neighbors. The resulting valence band is good and the conduction band very reasonable.

In the past there have been several attempts to model the energy bands of Si by a simple TB Hamiltonian and the need for TB total-energy representations to provide inter-atomic

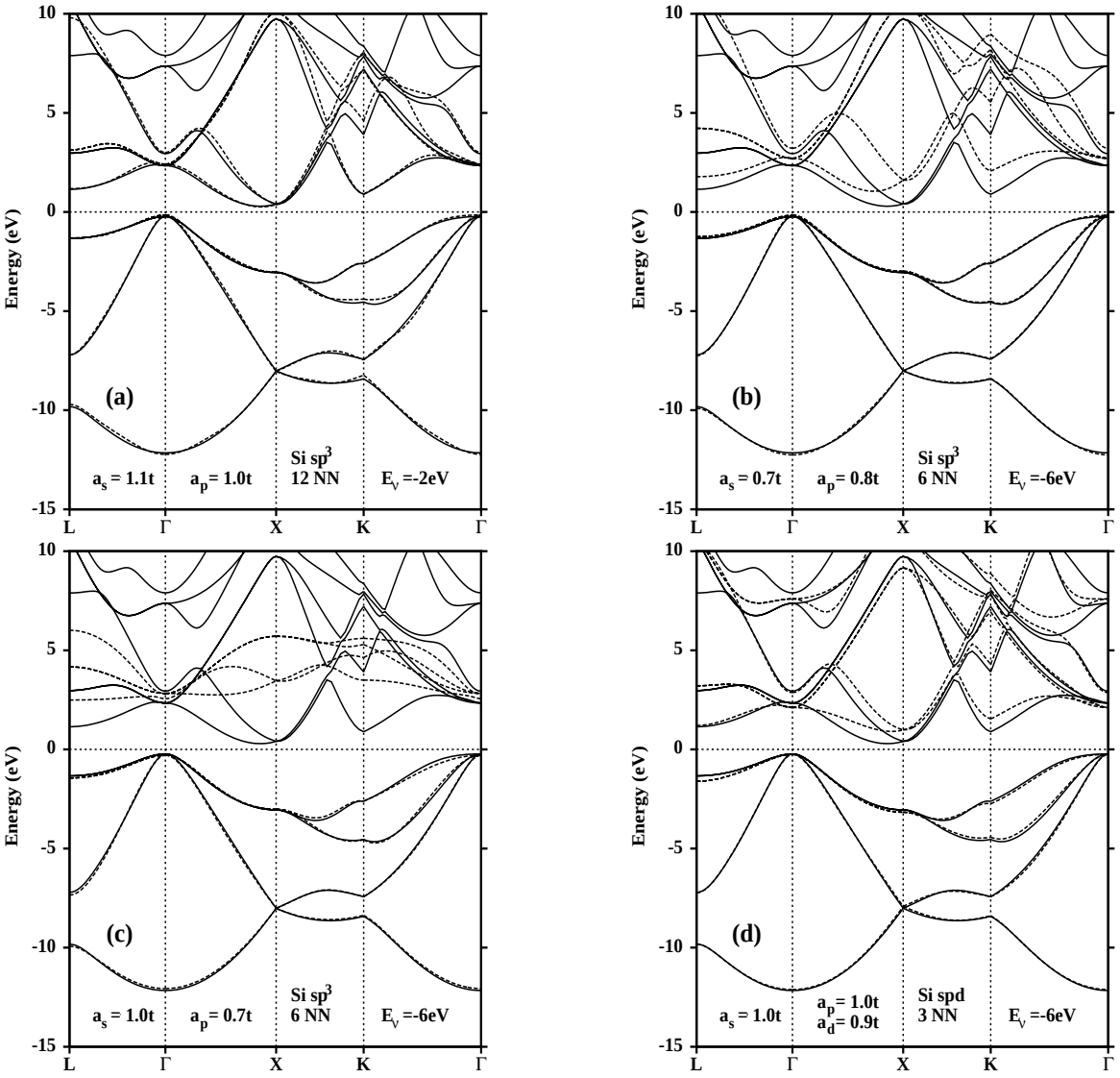


Figure 9: LDA energy bands of diamond structured Si in full lines and various TB approximations thereto in dashed lines (a)-(d). The corresponding LMTO sets and the real-space truncation of the Hamiltonian and overlap matrices (31) and (30) in (a) and (b), and of $h$ (33) in (c) and (d), are specified at the bottom of the panels.

forces for molecular-dynamics simulations has renewed this interest. These attempts range from simple nearest-neighbor, orthogonal parametrization of diamond structured Si with an $sp^3$ basis in the 70's[27] to recent work with long-ranged non-orthogonal $sp^3d^5$ basis sets[28] with a hope to provide transferable parameters. All these works relied on fittings of energy bands and total energies obtained from first-principles calculations. Our method is free from such fitting procedures and is purely deterministic. The recent work of McMahan and Klepeis[29] is more similar in spirit to ours, but being based on a full-potential multiple-kappa LMTO calculation with the need for subsequent contraction to a minimal $sp^3d^5$ basis set, it is more complicated and computationally far more demanding. In fact, our method is so fast, that for us, transferability is no issue. But in all fairness, our total-energy and force calculation is still pending.

GETTING RID OF THE EMPTY SPHERES

The full LDA potential for diamond-structured Si is shown in the top left of Fig. 10. What was used in the LMTO calculations of Fig. 9, however, was the conventional ASA potential shown in the top-right panel of Fig. 10, which is slightly overlapping [$\omega$=14%; see Eq. (24)] and, in addition to the Si-wells, has repulsive wells at the E-sites to describe the hills of the potential. Despite its crude appearance, this ASA SiE-potential, gives nearly exact LDA

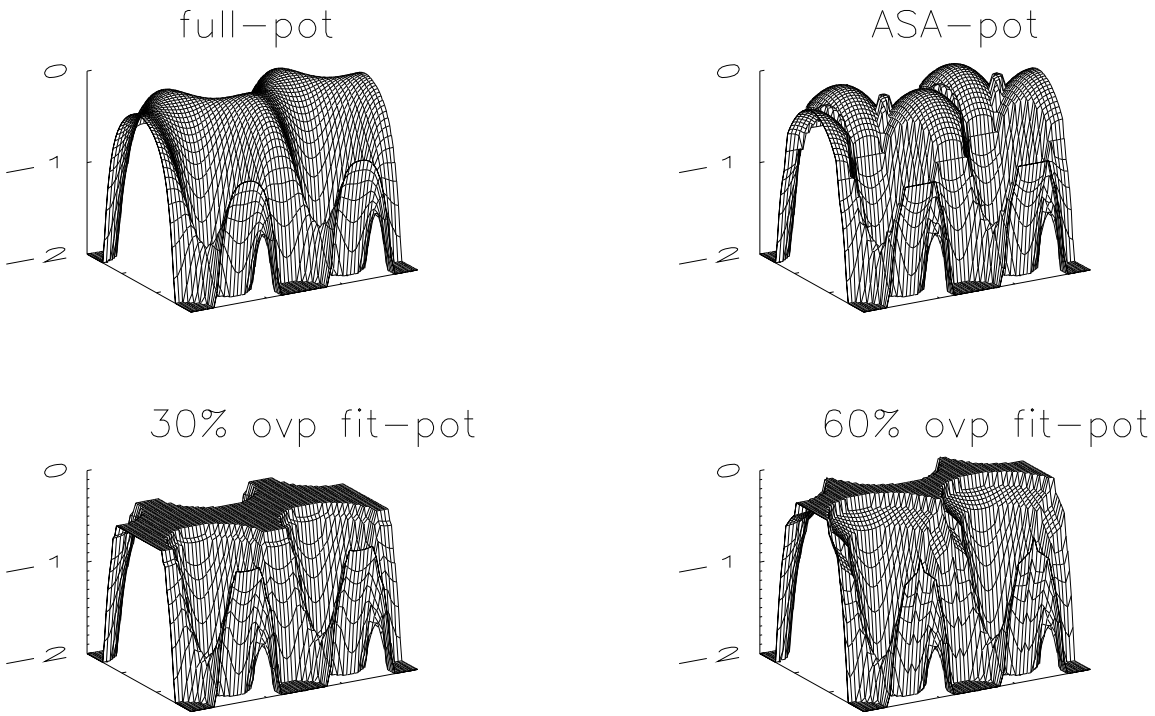


Figure 10: Full LDA potential (in Ry) and various MT approximations for diamond-structured Si in the (110)-plane. The two pseudo potentials at the bottom are least-squares fits to the ASA potential.

valence and conduction bands. But this is a special case. In general, the potential consists of spherically symmetric craters with hills in between, and the latter can be of any shape. Such a potential is naturally modelled by a superposition of atom-centered spherically-symmetric wells, and since we have proved in Eq. (20) that the KKR method can handle such a potential, unless the overlap is too large. The questions are whether this holds also for the new LMTO method, and whether the overlap allowed by these methods is sufficiently large that the MT zero moves up close to the hill tops and the wave functions tail properly off into the voids. Non-MT perturbations would then be local and simple to include. Therefore we first try to treat diamond-structured Si. Two appropriate potentials with respectively 30% and 60% radial overlap are shown in the bottom panels of Fig. 10.

We thus want to fit the full potential, $\mathsf{V}(\mathbf{r})$, to a constant (the MT zero) plus a superposition of spherical wells: $\mathsf{V}(\mathbf{r}) \sim V_{mtz} + \sum v_R(r_R) \equiv V(\mathbf{r})$. If we decide on a least-squares fit, that is, minimization of $[\mathsf{V} - V]^2$, then variation of the functions $v_R(r_R)$ leads to a set of coupled integral equations, one for each $R$ saying that the *spherical* average around site $R$ for radius $r_R$ should be the same for the sum of the MT wells as for the full potential minus $V_{mtz}$. Variation of $V_{mtz}$ leads to one equation saying that the average of the MT and the full potential should be the same. These equations are fairly simple to solve numerically, but they do not quite express what we want, because a volume element in a region like a void, where the electron has little chance of being, enters with the same weight in the fitting as a volume element in say the bond region. What we really want is a *pseudo* potential which, for a certain band, say the valence band, minimizes the mean squared deviation of the one-electron energies, $\mathrm{Tr}\rho[\mathsf{H} - H]^2 = \mathrm{Tr}\rho[\mathsf{V} - V]^2$, and this then brings in the electron density, $\rho(\mathbf{r})$, as weighting function. This weighting presents little problem for the $\delta V_{mtz}$-equation, which is merely: $\int [\mathsf{V} - V]\,\rho d^3r = 0$, but it complicates the $\delta v_R(r_R)$-equations so much, that we decided on keeping $\rho$ in the $\delta V_{mtz}$-equation only. Our MT pseudo potential[30] thus pseudizes the hills rather than the core regions.

Since at this stage, we merely want to see whether we can get rid of the empty spheres

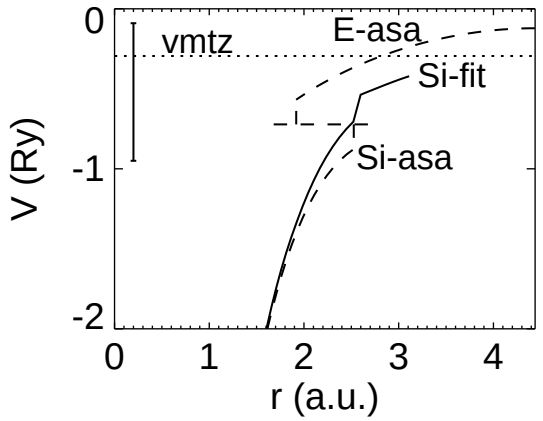


Figure 11: Radial behavior of the Si and E potential wells of the ASA potential (dashed), and of the Si well of the pseudo potential with 40% radial overlap (full). The corresponding 3-dimensional potentials are shown in the upper right and lower left of Fig. 10. The dotted line is the MT zero of the pseudo potential. The vertical bar at the left-hand side indicates the position and extent of the valence band.

in the diamond structure by comparing the valence-band structure calculated for the true potential with that calculated for its pseudo potential, we take the true potential to be one for which we can solve Schrödinger's equation with high accuracy, namely the ASA potential shown in the upper right of Fig. 10. Since this potential is discontinuous at the surfaces of the Si and E spheres, its pseudo potentials, shown at the bottom, are not only discontinuous at $s$, but also at the Si AS radius. The radial behaviors of the Si and E wells of the ASA potential, as well as that of the Si pseudo potential with 40% radial overlap, are shown in Fig. 11. By comparison of the pseudo potentials with 30% and 60% radial overlap shown at the bottom of Fig. 10, it is obvious that the latter resembles the true potential most closely. Whereas the MT zero of the 14% overlapping ASA potential is only slightly above the bottom of the valence band, that of the 40% overlapping pseudo potential lies 6 eV higher, and that of the 60% overlapping potential is at the top of the valence band.

We have now used the new LMTO method [Eq.s (30) and (31) with a Si $sp^3d^5$ LMTO set and the Si $f$-channels downfolded] to calculate the energy bands for the valence-band pseudo potentials as a function of the radial overlap $\omega$. The rms and mean errors of the calculated valence bands are shown in Fig. 12 by diamonds. Since for increasing $\omega$, the potential has increasing range and, hence, increasing freedom, the *rms* error initially falls, but it eventually rises again as the kinetic-energy errors given by Eq. (20) and proportional to $\omega^4$ take over. The minimum rms error of 80 meV per electron is reached at 30% overlap. The *mean* error we had expected to vanish for overlaps so small that the kinetic-energy errors are negligible, because the pseudo potential was constructed such that $\int [\mathsf{V} - V]\, \rho d^3r = 0$. Nevertheless, the computation yields a "background" mean error of –50 meV per electron. This is most likely due to errors of second order in $\mathsf{V} - V$ caused by the unphysical discontinuities at the E-spheres of the ASA potential. We expect this background error to vanish and thereby the

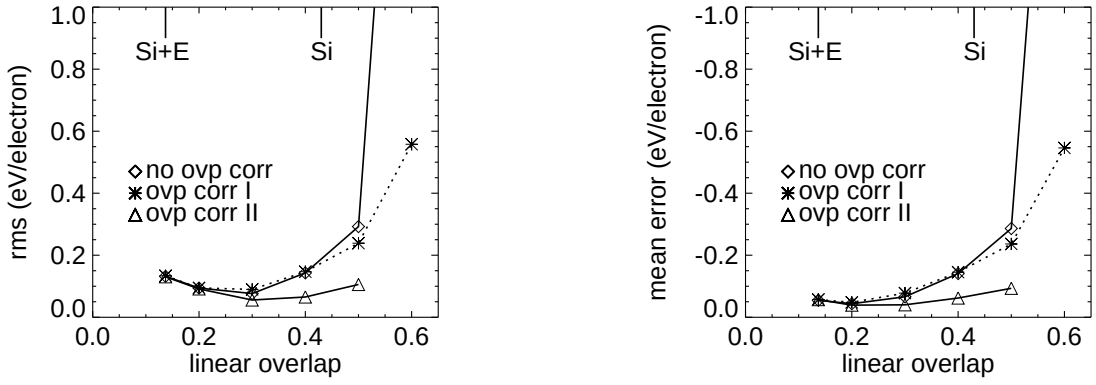


Figure 12: Rms and mean errors of the valence-band energies arising by pseudizing of the ASA potential of diamond structured Si in order to get rid of the E wells. The abscissa is the radial overlap, $\omega$, as defined in Eq. (24). For *space-filling* Si+E spheres, $\omega$=14%, and for Si spheres 43%. Overlap correction I modifies the pseudo potential, while II includes the proper kinetic energy in the LMTO Hamiltonian.

rms error to be reduced, when for $\mathsf{V}$ we use the full potential in the top left panel of Fig. 10.

Since the kinetic-energy error is negative, it represents an attraction between overlapping atoms, and this might cause problems in molecular-dynamics calculations. However, although this attraction increases rapidly with overlap, it does decrease for decreasing interatomic distance and fixed $s$-radii [see Eq. (23)].

If radial overlaps in excess of $\sim$30 % are needed, then the kinetic-energy error must be corrected. We have tried two schemes, the results of which are given in Fig. 12 by the stars and the triangles. In the first scheme (stars) we have merely modified the pseudo potentials by including in the $\delta V_{mtz}$-equation the kinetic-energy error to leading order as given by Eq. (23), whereby this equation becomes: $\int [\mathsf{V} - V]\, \rho d^3r = \sum_i \Delta\varepsilon_i$. This leads to a reduction of the overlap error, mainly through reduction of the discontinuity $v\,(s)$. This correction is very simple, but as seen from the figure, hardly sufficient because it only treats the error proportional to $v\,(s)^2\,\omega^4$. Our work on the second scheme (triangles) is still in progress.[30] Here, we evaluate the LMTO Hamiltonian matrix properly to all orders in the overlap, that is, we calculate the LMTO matrix elements following Eq. (20). Of course, this adds terms to expression (31) for the Hamiltonian and spoils the beauty of Eq.s (8), (33), and (34), but we wish to prove that we can control the overlap errors of the new LMTO method, and we want to investigate how large overlaps we can handle. The preliminary results shown in Fig. 12 are encouraging.

REFERENCES


[1] J. Korringa, Physica **13,** 392 (1947); W. Kohn and J. Rostoker, Phys. Rev. **94,** 1111 (1954).

[2] O.K. Andersen, A.V. Postnikov, and S. Yu. Savrasov, in *Applications of Multiple Scattering Theory to Materials Science,* eds. W.H. Butler, P.H. Dederichs, A. Gonis, and R.L. Weaver, MRS Symposia Proceedings No. 253 (Materials Research Society, Pittsburgh, 1992) p 37.

[3] O.K. Andersen, O. Jepsen, and G. Krier in *Lectures on Methods of Electronic Structure Calculations,* edited by V. Kumar, O.K. Andersen, and A. Mookerjee (World Scientific Publishing Co., Singapore, 1994), pp. 63-124.

[4] R. Zeller, P.H. Dederichs, B. Ujfalussy, L. Szunyogh, and P. Weinberger, Phys. Rev. B **52**, 8807 (1995).

[5] O.K. Andersen, Solid State Commun. **13**, 133 (1973); O.K. Andersen, Phys. Rev. B **12**, 3060 (1975); O. Jepsen, O.K. Andersen, and A.R. Mackintosh, Phys. Rev. **12**, 3084 (1975).

[6] O.K. Andersen and O. Jepsen, Phys. Rev. Lett. **53**, 2571 (1984).

[7] H.L. Skriver, *The LMTO Method* (Springer-Verlag, Berlin, 1984).

[8] H.L. Skriver and N.M. Rosengaard, Phys. Rev. B **43**, 9538 (1991); I. Turek, V. Drchal, J. Kudrnovsky, M. Sob, and P. Weinberger, *Electronic Structure of Disordered Alloys, Surfaces, and Interfaces* (Kluwer Academic Publishers, Boston/London/Dordrecht, 1997).

[9] O.K. Andersen, O. Jepsen and M. Sob, in *Lecture Notes in Physics: Electronic Band Structure and Its Applications*, eds. M. Yussouff (Springer-Verlag, Berlin, 1987).

[10]W.R.L. Lambrecht and O.K. Andersen, Phys. Rev. B **34**, 2439 (1986).

[11]O.K. Andersen, Z. Pawlowska and O. Jepsen, Phys. Rev. B **34**, 5253 (1986).

[12]D.G. Pettifor, J. Phys. F **7**, 613 (1977); *ibid.* **7**, 1009 (1977); *ibid.* **8**, 219 (1978).

[13]O.K. Andersen and R.G. Woolley, Mol. Phys. **26**, 905 (1973).

[14]R. Haydock in *Solid State Physics* **35** edited by H. Ehrenreich, F. Seitz, and D. Turnbull (Springer Verlag, Berlin, 1980) p. 129.

[15]H.J. Nowak, O.K. Andersen, T. Fujiwara, O. Jepsen and P. Vargas, Phys. Rev. B **44**, 3577 (1991); P. Vargas C. in *Lectures on Methods of Electronic Structure Calculations,* edited by V. Kumar, O.K. Andersen, and A. Mookerjee (World Scientific Publishing Co., Singapore, 1994), pp. 147-191; S. Frota-Pessoa, Phys. Rev. B **36**, 904 (1987); S.K. Bose, O. Jepsen, and O.K. Andersen, Phys. Rev. B **48**, 4265 (1993).

[16]The Stuttgart TB-LMTO program. http://www.mpi-stuttgart.mpg.de

[17]O. Jepsen and O.K. Andersen, Z. Phys. B **97**, 35 (1995).

[18]O. Gunnarsson, J. Harris, and R.O. Jones, Phys. Rev. B **15**, 3027 (1977); K.H. Weyrich, Solid State Commun. **54**, 975 (1985); M. Springborg and O.K. Andersen, J. Chem. Phys. **87**, 7125 (1986); M. Methfessel, Phys. Rev. **38**, 1537 (1988); M. Methfessel, C.O. Rodriguez, and O.K. Andersen, Phys. Rev. B **40**, 2009 (1989); J. Wills (unpublished); S.Y. Savrasov, Phys. Rev. B **54**, 16470 (1996).

[19]An exception is: L. Vitos, J. Kollar, and H.L. Skriver, Phys. Rev. B **49**, 16694 (1994).

[20]R.W. Tank, O. K. Andersen, G. Krier, C. Arcangeli, and O. Jepsen (unpublished).

[21]In order to make $\dot{K}$ positive-, rather than negative definite, we have defined $K$ with the opposite sign as in Ref. 3.

[22]This neglects the high-order term $\langle \phi - \varphi \mid \psi - \varphi \rangle + h.c. \sim \frac{1}{24} s \left(s-a\right)^4 v\left(s\right) \phi\left(s\right) K$.

[23]O. K. Andersen, O. Jepsen, A. I. Liechtenstein, and I. I. Mazin; Phys. Rev. B **49**, 4145 (1994); O. K. Andersen, A. I. Liechtenstein, O. Jepsen, and F. Paulsen; J. Phys. Chem. Solids **56**, 1573 (1995).

[24]T. Saha-Dasgupta, O. K. Andersen, G. Krier, C. Arcangeli, R.W. Tank, O. Jepsen, and I. Dasgupta (unpublished).

[25]T. F. A. Müller, V. Anisimov, T. M. Rice, I. Dasgupta, and T. Saha-Dasgupta (unpublished) cond-mat/9802029.

[26]S. Y. Savrasov and O. K. Andersen, Phys. Rev. Lett. **77**, 4430 (1996).

[27]D. J. Chadi and M. L. Cohen, phys. status solidi **68**, 405 (1975).

[28]R. E. Cohen, L. Stixrude, and E. Wasserman, Phys. Rev. B **56**, 8575 (1997).

[29]A. K. McMahan and J. E. Klepeis, Phys. Rev. B **56**, 12 250 (1997).

[30]C. Arcangeli, O.K. Andersen, and R.W. Tank (unpublished).